"Something's Gotta Give": Rethinking linguistic models of rhythm and text-setting through evidence from jazz bop swing.

Stephanie Sin-yun Shih

This thesis satisfies the requirements of the H195 sequence for Honors in the Bachelors Degree of Linguistics.

University of California, Berkeley
May 2007

This thesis by Stephanie Sin-yun Shih, entitled " 'Something's Gotta Give': Rethinking linguistic models of rhythm and text-setting in jazz bop swing," satisfies the requirements of the H195 sequence for Honors in the Bachelors Degree of Linguistics.

Kristin Hanson, Co-advisor

Sharon Inkelas, Co-advisor

University of California, Berkeley
May 2007

“Something’s Gotta Give”: Rethinking linguistic models of rhythm and text-setting through evidence from jazz bop swing.

by Stephanie Sin-yun Shih

Contents

Abstract

Jackendoff and Lerdahl's (1983) grid-based theory of modeling musical intuitions of rhythm—based on Western classical music—has formed the foundation for current research on musical rhythm and text-setting, which has also primarily focused on classical music. Using unanalyzed data from Ella Fitzgerald's 1950-1960s recordings, this paper reevaluates the previous linguistic models of rhythm and text-setting with non-classical evidence. The conclusion is two-fold.

Given the distinct nature of jazz "swing" rhythm, changes must be made in the grid-based model of musical rhythm. Based on evidence from the separate roles of drums, bass, and melodic instruments, I generate a metrical model of jazz rhythm that departs from Jackendoff and Lerdahl's (1983) in three significant ways. First, the polyphonic texture of jazz necessitates the existence of two metrical models which together capture jazz's rhythmic form, though these models are similar in many ways and may be reconciled as one. Second, evidence from jazz supports an argument that musical rhythm, like linguistic rhythm, involves the metrical features of culminativity and constituency. This argument leads to the use of a tree-based model for representing rhythm as opposed to the grid-based model, which fails to illustrate constituency structure. Third, unlike the binary metrical model hitherto used to describe musical rhythm, jazz rhythm requires ternary metrical structure, like anapestic meters, accounting for both the seemingly uneven rhythm of swing and jazz's conflict between stressing phrasal beginnings and metrically prominent beats. This model of jazz rhythm describes more of the metrical relationships present in both music and language than possible in Jackendoff and Lerdahl's (1983) model, allowing for more flexibility and comprehensiveness in describing listeners' intuitions of musical rhythm.

The revised model of musical rhythm necessitates re-consideration of a process that relies heavily on rhythm—text-setting. The tandem and conflict-based model of jazz rhythm supports the recent arguments (i.e., Hayes 1997, 2005, Kiparsky 2005, 2006) for Optimality Theory in the generative analysis of text-setting. This paper furthermore argues for a gradient Optimality-theoretic approach to economize the text-setting system and to mirror the relative nature of the tree-based rhythmic model presented here. These claims are demonstrated in the constraints and constraint ranking for text-setting in jazz bop swing.

Acknowledgements

First, to my advisors—Kristin Hanson and Sharon Inkelas. Without their constant support and encouragement, incredibly impressive insights far beyond my abilities, and stalwart patience for my slow progress, this thesis would not have happened. But their role in this is much larger than just that. Without Kristin Hanson, I would not be a linguist today; her inspiring English 179 "Literature and Linguistics" lectures on rhythm, meter, music, Shakespeare, rhyme, and Bob Dylan captured my musical imagination. Ever since that semester, she's been undeniably the driving force behind my linguistic pursuits. And without Sharon Inkelas, I would not be a phonologist today; her Ling 115 "Phonology and Morphology" and 211A "Advanced Phonological Theory" classes were like Tetris with phonological knowledge to me—everything fell and fit together so perfectly. I am the luckiest metrics-thesis-writer in the world to have these two professors volunteer to co-advise me and give so much of their time to worrying over my trees, grids, presentations, stress-matching constraints, the paper, and everything.

This research was assisted by a grant from the Summer Undergraduate Research Fellowship (SURF) Program in the College of Letters and Sciences at UC Berkeley during Summer 2006. This grant gave me my much needed data, since CDs these days are quite pricey. Moreover, acknowledgements once more to Kristin Hanson, who gave up precious summertime hours in her beautiful garden to advise me.

My gratitude as well to the audiences at the Berkeley linguistics department colloquium, Phonology Phorum, and the Society of Linguistic Undergraduate Students fifth annual symposium for their questions, comments, and posers. This department and school has been a wonderful place to be a fledgling linguist: faculty and staff, graduate students, undergraduates—

there’s always someone to bounce ideas off of and to make linguistic jokes to walking down the long Dwinelle hallway.

Finally, to my family and friends and all of my music teachers for everything they’ve done over the years, from their musical fellowship to their friendship and their constant *ganbatte*'s.

0. Introduction

Jazz has long been considered America's own form of Western music, distinguished from the older European classical musical tradition in part by its "swing," a syncopated and perceptibly uneven rhythm. Syncopation is defined as "rhythmic displacement" (Macy 2006), caused by accentuation of beats that the listener does not expect to be accentuated, and this frequently occurs in jazz, resulting in the distinct rhythmic character of swing. Given its variation from other Western musical traditions, the underlying rhythmic properties of swing have not been as widely explored and formalized as those of Western classical tonal music—the same is true for aspects of swing that rely heavily on rhythm such as text-setting, putting together words and music.

Text-setting in English, at its most basic hypothesis, is the matching of linguistic stress to musical stress, the pairing of strong syllables in language to strong beats in music. It is in the exploration of the text-setting of English to jazz that an accurate description and model of swing rhythm, which has hitherto not been developed, is imperative; moreover, the issue of text-setting itself reflects a musical rhythmic structure, as it has been shown in folk verse that linguistic stress systematically concurs with musical stress (e.g., Halle and Lerdahl 1993, Hayes and Kaun 1996, Hayes 2005). In this paper, I develop a model of jazz rhythm and discuss issues of text-setting using both empirical data in Ella Fitzgerald recordings and overt instructional knowledge from jazz education in order to answer the question of the underlying properties of jazz swing rhythm.

My focus in this research is primarily on jazz bop swing, a mainstream jazz form from the 1950s-60s that is the most commonly recognized subgenre of jazz involving the rhythmic characteristic of swing. Jazz originated in America at the beginning of the twentieth century, arising out of an eclectic mix of piano ragtime, Dixieland marching band music, and Western classical music. The first major era of jazz was the "Swing Era" of the twenties and thirties, when jazz served as dancing music and was characterized by a simple harmonic structure but a complicated, syncopated rhythm—swing—different from that of Western classical music. Following World War II, a different form of jazz developed in the "Bebop Era" when jazz turned from dance music to listening music. Bebop involved more complex harmonic and rhythmic structures than swing, further increasing the role of syncopation and unevenness. In the following decades of the fifties and sixties, mainstream jazz—jazz bop swing—integrated elements of both pre-forties swing, forties bebop, and the later forms of hard bop and cool bop to form a representative amalgamation of bop and swing.

One of the most notable musicians of the mainstream bop swing era is Ella Fitzgerald, known as jazz's "First Lady of Song." Fitzgerald's career spanned the swing era in the 1930s and the bop swing era of the forties to last throughout the late eighties. Having influences from both swing and bop, Fitzgerald was hugely popular in the mid twentieth century, reaching the height of her career around the fifties and sixties' jazz bop swing. Along with regular albums with a variety of songs, such as *Ella Sings Sweet Songs for Swingers* (1958-59), Fitzgerald recorded a series of songbooks honoring noteworthy jazz composers during this time. These songbooks marked the development of the "Great American Songbook," a canon of famous pieces by quintessential American jazz writers including Cole Porter, Rodgers and Hart, Duke Ellington, Irving Berlin, George and Ira Gershwin, Harold Arlen, and Jerome Kern. In addition to these composer songbooks, Fitzgerald recorded one more honoring lyricist Johnny Mercer—the only lyricist Fitzgerald honored in her songbooks collection. This honor is a tribute to Mercers' distinction as a definitive jazz lyricist. For the empirical data in this research, I have chosen to use Fitzgerald's bop swing recordings of Johnny Mercer's songs, including her *Johnny Mercer Songbook* (1964), as both Fitzgerald and Mercer are exemplary figures of jazz bop swing.

This paper is organized as follows: section one discusses previous approaches to modeling musical rhythm; section two proposes a metrical model for jazz bop swing rhythm; section three discusses previous research on text-setting; and section four will discuss issues in bop swing text-setting in the framework of Optimality Theory.

1. Background: Musical rhythm

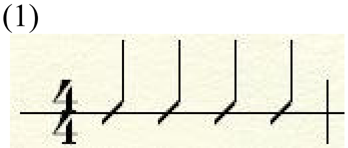
The traditional notation of music acts much like shorthand, encoding a guideline of the music, but neglecting to represent every detail, however important. Most notably, traditional musical notation does not really reflect any rhythmic structure; for example, when a musician reads a piece of written music, nothing tells him that the first beat of every measure should be the strongest, just as English orthography never depicts lexical stress. To deal with this problem, it is necessary to describe the properties and characteristics of music intuitive to an experienced listener—a person who has enough musical experience and intuition to be able to classify an unknown piece as exemplary of a certain musical idiom (Jackendoff and Lerdahl 1983)—similarly to describing the intuitions of native language speakers. Linguistic approaches to describing rhythm in music, spearheaded by Jackendoff and Lerdahl (1983) in *The Generative Theory of Tonal Music* (henceforth, *GTTM*), are summarized in this section.

1.1 Musical Meter

The most common musical meter in most Western classical music and jazz bop swing is four/four time, meaning that each measure—also known as a bar—is made up of four quarter notes, each worth one beat. A measure acts as a denotation of when primary stress repeats, and in four/four time, a bar means that some sort of primary stress will reoccur every four beats. Traditionally in Western classical notation, stress may be inferred as the beat immediately following a bar line; however, because jazz has a perceptibly different rhythmic structure than Western classical music, I will not proceed with this assumption. Even though there are many other time signatures in music, in this paper I only discuss the rhythms of jazz bop swing in four/four time, which is also known as common time. I avoid the other musical meters for two reasons: firstly, isolating a certain time signature will allow me to make definite claims about the metrical structure of swing without risking possible variations caused by natural subdivisions in measures, and secondly, the sheer rarity of pieces in other time signatures in the jazz bop swing era evidences certain privilege in status and importance of four/four time in the genre.

As for canonical musical notation, shown below in (1) is a measure of four/four time with quarter notes in standard notation:

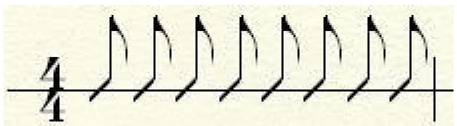
(1)



1 2 3 4

As represented by the bold numbers is each quarter note beat of the measure; this is also how the music is orally counted and communicated by musicians. If each quarter note were to be broken down into further subdivisions—eighth notes, for example—then they would be as demonstrated in (2):

(2)



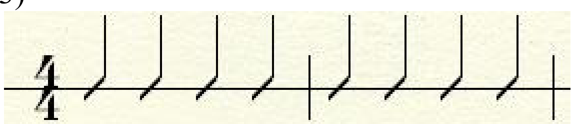
1 1+ 2 2+ 3 3+ 4 4+

Once again, the bold numbers represent how the counts are referred to.

Even though canonical musical notation acts more as a short-hand than a direct representation of everything happening in the music, the method used to "count" beats, shown in (1) and (2), reflects a musical intuition of a tactus level. The tactus level is the metrical level at which beats are most salient, described by Jackendoff and Lerdahl as "the level of beats… with which one most naturally coordinates foot-tapping or dance steps" (Jackendoff and Lerdahl 71). They furthermore lay out additional parameters for determining the tactus level of a piece: the tactus level will never leave the range of forty to 160 beats per minute, this level is not far away from the smallest metrical level of notes, and the tactus level remains constant throughout the entire piece of music (Jackendoff and Lerdahl 73-74). Additionally, harmonies will not change below tactus level, meaning that harmonic structure is somewhat dependent on rhythmic perception and the tactus. According to this description of the tactus level, the quarter note beats generally form the tactus level in a Western classical composition in four/four time, meaning that the quarter note level is the main metrical level at which rhythm is perceived. This intuition of tactus is registered in the canonical musical nomenclature, as quarter note beats of a four/four time measure receive whole-numbered counts and, when subdivided into eighth notes, the second eighth note receives an **"**+," not a whole-numbered count.

In jazz bop swing, the issue of a tactus level is slightly more difficult. Jackendoff and Lerdahl assert that the tactus level is where an experienced listener would tap his feet, and for Western classical four/four time music, this occurs on the quarter note level. For most four/four time swing experienced listeners may tap their feet on each quarter note—this approach is completely “grammatical”; however, more commonly, listeners tap their feet on beats **2** and **4** of the music—tapping only on **1** and **3** feels quite unnatural and unintuitive. But, it is also impossible to say that bop swing has units that begin on the second or fourth beats of a measure, for if the same experienced listener were to orally count the beats while tapping his foot, he would count **1** as the beginning of a bar while still tapping on **2** and **4**, shown in (3):

(3)



1 2 3 4 1 2 3 4

tap tap tap tap

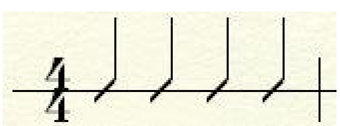
Moreover, if **2** were the beginning of the swing tactus level, then it naturally would take over the **1** beat position, as “**1**” is the beginning of a tactus unit; somehow, marking the beginning or boundary of the measure acts independently from marking prominence in the conventional practice of counting. I must therefore assume that the tactus level of common time bop swing is the quarter note beat, and note that tapping one’s foot on each beat at this level is also possible and does not sound unnatural or unintuitive. The instinctive tapping on beats **2** and **4** instead demarcates the more prominent beats of the tactus level, a point which I will discuss further later. The assertion of a quarter note metrical level as the tactus of jazz bop swing concurs with Jackendoff and Lerdahl’s parameters of a tactus level, generally never leaving the range of forty to 160 beats per minute in these recordings, never being far away from the smallest metrical level of notes, and remaining constant throughout each song.

1.2 Modeling musical meter

In addition to the tactus level, it is necessary to modeling musical rhythm and text-setting to know which beats are more prominent than others. I use “prominence” here to mean a beat which is relatively stronger or more salient than its counterparts. Prominence is different from the traditional musical notion of accent, represented by “<”. An accent (<) usually has an association with being aurally louder (Randel 3). In actuality, many factors may realize prominence and denote the strength of a beat: dynamic factors, meaning changes in volume; tonic factors, shifts in harmony; and agogic factors, subtle and slight differences in the duration of notes. To represent prominence—when there are accentuations of certain notes and beats—while using musical notation, I will use a “*” above each prominent note in my nomenclature, as demonstrated to avoid the inflexible notion of volume (<) as the sole indicator prominence.

Jackendoff and Lerdahl’s *GTTM* explores the patterns of prominence in Western classical music and formulate a model of musical rhythm. In a typical bar of Western classical music, the first and third quarter note beats are more prominent than the second and fourth beats. Additionally, between **1** and **3**, beat **1** will be the strongest overall (4):

```
(4)    *
       *     *
```



1 2 3 4

Jackendoff and Lerdahl translated the representation in (4) into a metrical model of Western classical music rhythm, shown in (5). The stronger beats at any given level are represented by being present on the immediate level above, demonstrating that they are more salient than other beats. Thusly, because beats **1** and **3** are stronger than **2** and **4**, they are present on the level above, and be because beat **1** is strongest overall, it is also represented at the highest level.

```
(5)    *           L(2)
       *     *     L(1)
       *  *  *  *  L(0)
       1  2  3  4
```

The grid in (5) is Jackendoff and Lerdahl's basic template of four-four time Western classical rhythm, which is a falling stress pattern, going from strong to weak.

The *GTTM* rhythmic model is accompanied by rules governing both the structure of the grid and the preference of other musical phenomena such as notes and chords to occur in each metrical position. Those rules governing the basic structure of the grid constitute the set of Metrical Well-formedness Rules (MWFR), most of which Jackendoff and Lerdahl claim to be universal (Jackendoff and Lerdahl 69). The MWFRs are reproduced in (6)[1]:

(6) MWFRs

MWFR1: Every attack point must be associated with a beat at the smallest metrical level present at that point in the piece.

MWFR2: Every beat at a given level must also be a beat at all smaller levels present at that point in the piece.

MWFR3: At each metrical level, strong beats are spaced either two or three beats apart.

MWFR4: The tactus and immediately larger metrical levels must consist of beats equally spaced throughout the piece. At subtactus metrical levels, weak beats must be equally spaced between the surrounding strong beats.

(347 Jackendoff and Lerdahl 1983)

MWFR1 and MWFR2 are Jackendoff and Lerdahl's universal Metrical Well-Formedness Rules. MWFR1 restricts the generation of a grid (7) that represents beats that are not beats at the lowest level.

(7) Outlawed by MWFR1

```
  *       *
*    *    *    *
```

This universal rule must apply in any model of meter, since a beat cannot be prominent if it does not exist at all. This does not mean that a note must be played in the beat position, however, because the assignment of notes is only a manifestation of the underlying rhythmic structure, not

[1] Jackendoff and Lerdahl's definition of a "beat" is each star in the grid, whereas I use "beat" to refer to the note numbers in (1) and (2).

a direct reflection of it. MWFR2 is a rule analogous to the Continuous Column Constraint (CCC) (34 Hayes 1995) for linguistic rhythm, which formalizes the ungrammaticality of the structure in (8):

(8) Outlawed by MWFR2 and CCC

```
        *
*       *
*   *   *   *
```

Similar to MWFR1, MWFR2 asserts that, in grid representations, it is impossible for something that was not prominent on all lower levels to be prominent on higher levels, since the very definition of strength in grid modeling is that the strongest beat dominates the column with the greatest number of stars (*). By not having a * on any given line means that the beat is not prominent in comparison with other prominent beats and because it is already weaker, cannot be stronger at higher levels. The Metrical Well-Formedness Rules not only constrain theoretically the type of grid that can be generated but also reflect the behavior of prominence in Western classical tonal music.

In addition to the MWFRs, Jackendoff and Lerdahl assert some universal rules for mapping musical events to the metrical grid in the form of Metrical Preference Rules (MPR). The preference rules which will be most relevant to this paper are listed in (9):

(9) MPRs

MPR3: Prefer a metrical structure in which beats of level L_i that coincide with the inception of pitch-events are strong beats of L_i.

MPR4: Prefer a metrical structure in which beats of level L_i that are stressed[2] are strong beats of L_i.

MPR5: Prefer a metrical structure in which a relatively strong beat occurs at the inception of relatively long durations of pitch events, dynamics, motif, phrase, harmony, etc.

(347-348 Jackendoff and Lerdahl 1983)

[2] Jackendoff and Lerdahl use "stress" here to mean "accent" (<).

For the most part, these MPR may be collapsed into one preference rule: dynamic, tonic, and agogic events all prefer to begin on prominent beats. In terms of the rhythmic grid, the MPR outlines the preference for accentuated events to occur on beats with more stars (*) than others.

The rhythmic grid, along with its rules that govern both the structure of the grid and the preferred association of musical events, models only the rhythmic structure of music. For modeling the syntax-like structure of higher level phrasal constituents such as harmony, Jackendoff and Lerdahl propose a time-span reduction model, which functions much like a syntactic tree in languages. The time-span reduction tree represents musical phrases that can span across the entire length of a piece while the grid model describes more local phenomena. Because text-setting is more of an issue of phonological properties, I will only discuss the rhythmic modeling of music in this paper.

All of the rhythmic properties of music described in this section are demonstrated by Jackendoff and Lerdahl to apply to Western classical music and are hypothesized to be universal. But, because the surface rhythms of jazz bop swing differ a good deal from those of Western classical music, I explore in §2 the limitations of grid notation in describing the underlying rhythm of bop swing and discuss modifications that must be made to the metrical structure in order to accommodate different rhythmic properties of swing.

2. The basic rhythmic model of bop swing

Because the rhythm of jazz is perceivably disparate from the rhythm of Western classical music, a metrical model of bop swing is developed in this section in order to explore text-setting to swing rhythm in later sections. The data here has been transcribed from recordings, since jazz written composition does not fully describe heard musical events. In jazz, a written composition is a lead-sheet. The lead-sheet consists only of a melody, possible lyrics, and chord symbols; thusly, not much of the music heard in jazz is produced from only reading the written music in front of the musician. For example, comparing the lead-sheet of "That Ol' Black Magic" in (10) to Fitzgerald's 1960 recording, the page does not represent so much of the music: the bass, the drums, the piano—even Fitzgerald's interpretation of the lyrics, rhythms, and notes hardly matches to the written score.

(10)

In this point, jazz is quite different from most Western classical music. Though written classical music may not show every musical intuition in ink, most of what is heard is represented on the written page. It is insufficient to account for this difference in jazz and classical by attributing jazz to being an extemporaneous art form, that whatever a musician desires to play, he can. There must be constraints that identify the genre of jazz bop swing. The question becomes, therefore, "If the written notation does not represent all of the actual produced and heard music, then what does the experienced listener know that identifies the music as jazz?"

One most obvious part in jazz not written into the lead-sheet score is the rhythm section's role. The rhythm section, for my purposes, comprises of the bass and drums because these instruments have rhythmic roles that maintain a nearly constant rhythmic structure and pattern across almost every song, though these patterns may vary a little from song to song, but I account for this variation later. I show that the steady patterns of the rhythm section instruments (henceforth r-s instruments) form a regular, rising metrical model much like an iambic meter in poetics that can be represented by a simple grid template; moreover, the bop swing model of metrical strength dictated by the r-s instruments behaves in similar ways as poetic iambic meter, particularly in relation to boundary marking.

But, taking into account the rhythmic patterns of the harmonic and melodic instruments (henceforth h-m instruments), such as the vocalist, saxophonist, pianist, and so on, complicates the metrical model based on the r-s instruments. Because the stress patterns of the h-m instruments are more complex than those of the r-s instruments, a simple rising-rhythm metrical template cannot sufficiently or accurately represent the swing rhythms of these instruments. I therefore argue for an adaptation of the metrical tree model to account for these complications instead of the grid used for the rhythm section parts.

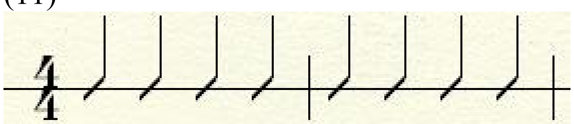
Lastly, I attempt to reconcile these two differing metrical models for the rhythm and the h-m instruments, since both metrical templates, though each somewhat distinctive, share several parallels. I assert that these two rhythmic roles in jazz bop swing can be collapsed into a metrical form much like that of anapestic meters, which can account for the complications posed by both separate models and its representations of the patterns of the rhythm section and the harmonic, melodic section.

2.1 The rhythm section (r-s)

The role of the bass and drums is often described by musicians as denoting the tempo of a piece, which is how fast the beats pass—hence, the name given to the unit of instruments, the "rhythm section." Due to this paramount role of marking "time" in a piece, the bass and drums nearly always play regular periodic rhythmic patterns, so that the identification of musical time facilitates the perception of the piece. Traditionally, the jazz rhythm section includes drums, bass, and piano; however, I have left the piano out of my description of the rhythm section as it does not serve any regularly cyclical rhythmic role as the bass and drums do. In this section, I develop a metrical model that governs these rhythmic patterns of the bass and drums because these are the instruments that are most clear with the bop swing rhythm.

The bass nearly always plays a pattern known as a "walking bass line," meaning that it strikes a sound on each quarter note—each beat of the tactus level, thus outlining the importance of the quarter beats as the tactus level. In addition to the bass's role of constant quarter notes, it stresses beats **2** and **4** in each measure, creating a weak-strong, rising pattern in the quarter notes:

(11) * * * *



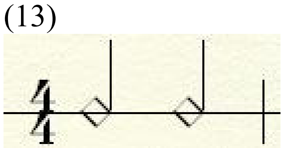
1 2 3 4 1 2 3 4

Examples of this pattern can be heard in "That Ol' Black Magic" and "Something's Gotta Give." This thus far forms a metrical template for bop swing rhythm as shown in (12):

(12)					
		*		*	L(1)
	*	*	*	*	L(0)
	1	2	3	4	

Sometimes, however, the bass does not play four quarter notes per measure but rather two half notes—striking once on beat **1** and once on beat **3** (13), as heard in pieces such as "Too Marvelous for Words" and "Day In, Day Out."

(13)

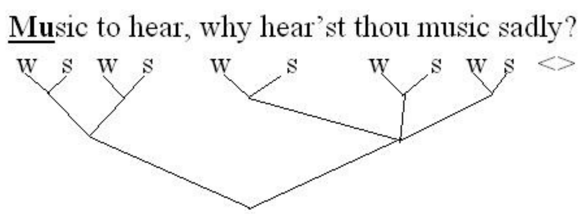


1 2 3 4

In pieces where this rhythmic pattern occurs, the bass will often alternate between the rhythmic pattern in (1) and the one in (3). This alternation and ability to switch between each pattern offers evidence for an unaltered metrical structure and no deviation from the metrical grid in (12). Because the tempo—the speed of the piece—is constant, there should also therefore be no change in the tactus level. Additionally, an experienced listener would still tap his foot or snap his fingers on the second and fourth beats despite the bass playing on **1** and **3**, demonstrating that no changes occur in tactus level or beat-prominence.

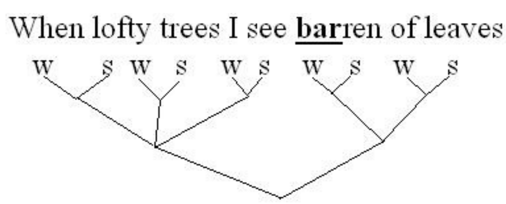
These explanations still do not answer the question of why the bass can play a regular bass line by striking notes only on beats **1** and **3** without also playing on the hitherto proved prominent **2** and **4**. Perhaps, then beats **1** and **3** have a different sort of strength. Related are iambic meters, in which we often find inversions of strength. Jespersen (1933) notes that these inversions occur at line-initially, and others after him (i.e., Hayes 1989) have noted that inversions may also occur at the beginnings of phonological phrases. Hanson (2003) shows that the occurrence of these inversions at the beginnings of lines and cola boundaries indicates a pressure for signaling beginnings, demonstrated in (14):

(14) (a)



(8.1)

(b)



(12.5)

(Shakespeare Sonnets)

In (14a), inversion occurs at the beginning of the line, and in (14b), inversion occurs at the beginning of a high-level constituent, though not at the beginning. The opposite kind of inversion does not happen in trochaic meters where the beginnings of higher-level constituents are already stronger than their counterparts. Most importantly, this inversion exists simultaneously with the weak-strong pattern of iambic pentameter without disrupting the meter. The bass's rhythmic role similarly mirrors this desire to demarcate beginnings; consequently, its obligatory pattern of playing on **1** and **3** delineates and emphasizes the left-most boundaries of the half-measure and the measure, of which **2** and **4** remain the prominent beats. The strength that a boundary beginning receives should not disrupt or change metrical structure since a stressed beginning is inherently part of the natural meter.

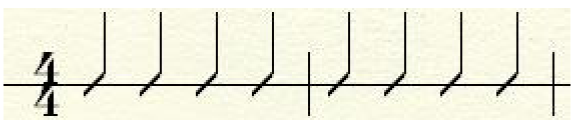
Another complication in the basic rhythmic structure modeled in (12) emerges in the bass's harmonic role. Most often, the bass will play different pitches for each tactus beat, but in the instances where pitches are repeated, the first two quarter notes will have the same pitch, and a change of pitch will occur starting on the third quarter note ("That Ol' Black Magic"). The bass's harmonic role is to "outline"—define—the harmonic changes written in the music, and these changes nearly always occur on beats **1** and **3** of a measure, just as written on the lead-sheet. This harmonic information, according to *GTTM*, offers counterevidence to the weak-strong pattern outlined above. Jackendoff and Lerdahl's Metrical Preference Rule 5 claims that harmonic change has a tendency of occurring on the strongest beats (Jackendoff and Lerdahl 84). Given my analysis above, however, these harmonic changes in bop swing—instead of demonstrating metrical prominence as Jackendoff and Lerdahl assert—demonstrate the signaling of left boundaries, just like the role that inversion plays in iambic meters. Based on just the rhythmic patterns of the bass in the rhythm section, we can formulate a metrical template for jazz bop swing as shown in (12). This metrical structure of swing closely parallels aspects of iambic meters, including the rising stress pattern and the propensity of marking or signaling beginnings.

The rhythmic pattern of the drums consists of several parts, as many instruments—usually played by one person[3]—make up one gestalt sound. Each of the parts of the drums' rhythmic pattern plays a role in denoting time, either outlining the tactus level or accentuating stressed beats. Like the bass, the drums' patterns identify beats **2** and **4** as the strongest in a measure; unlike the bass, the drums also offer evidence for a higher level of prominence, making beat **4** the strongest throughout the entire measure.

The parts of the drums' role is outlined in (15):

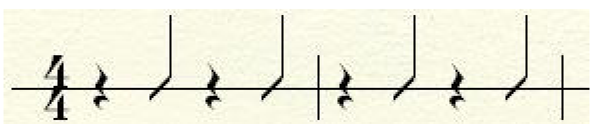
(15) (a) Bass drum (optional)

* * * *



1 2 3 4 1 2 3 4

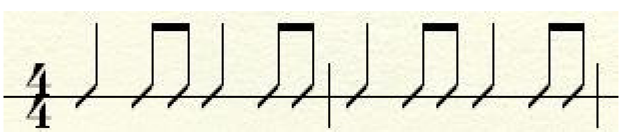
(b) Hi-hat cymbals



1 2 3 4 1 2 3 4

(c) Cymbals (either on the ride or elsewhere)

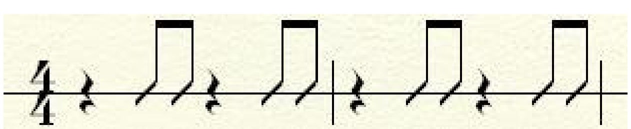
* * * *



1 2 + 3 4 + 1 2 + 3 4 +

(d) Other (snare drum or tom-toms)

* * * *



1 2 + 3 4 + 1 2 + 3 4 +

[3] Exceptions occur in some Latin and funk styles.

Examples of these patterns can be heard throughout every song, although the parts of the drum kit that each rhythm will occur on may vary from piece to piece. (15a) is the only pattern that is more difficult to hear, due to the low volume of the bass drum and its lower frequency which does not come across on most recordings—especially ones of the fifties and sixties. This bass-drum pattern is important to the music, though, because jazz drummers teach it as an integral part of bop swing drumming. (15b) illustrates that the hi-hat cymbals strike on beats **2** and **4**. This concurs with the metrical structure of the bass presented in (12) because the high hat is accentuating the beats as a part of the overall rhythms in the drums.

Illustrations (15c) and (15d) are different, however, because these patterns involve a subtactus level that has not yet been discussed in the structure of jazz bop swing rhythm. Here, the second and fourth beats are subdivided into two eighth notes while beats **1** and **3** remain unaltered. Setting aside the long-short distinction of these eighth notes, the question is whether this splitting of beats suggests a weakening of the beats, leaving **1** and **3** stronger. In the cymbal and snare pattern, **2** and **4** are still accented, just as in the bass and bass drum rhythmic patterns; consequently, based solely on this information, there is no reason to alter the metrical structure as presented in (12). As for the extra subdivision, the eighth note can actually be an indication of strength rather than weakness. The "filling in" with notes of the second and fourth beats emphasizes their prominence by adding more distinction to the beats. This subdivision may also act in the same way that extrametricality does in meter, as suggested by Hanson (2003), demarcating the right-most boundaries of constituents, similar to how inversion signals left-most boundaries. The drum patterns therefore fit with the metrical structure of (12), offering further evidence for the strength of beats **2** and **4**.

At a higher musical phrasal structure level, the drums often play a fill—defined as “a short harmonic, rhythmic, or melodic figure…[that usually] lasts no more than a beat or two” (Witmer 2006)[4]—at the end of phrasal constituents, often two, four, eight, or twelve bars long. Although I am not concerned with musical structure beyond the immediate metrical level, these higher-level fills are realized in ways that are relevant to prominences in metrical structure. These fills mark the right-most boundaries of phrasal constituents beyond the metrical level, analogous to the subdivisions of beats **2** and **4** in the lower-level drum patterns that signal endings of metrical constituents. Additionally, these phrasal level drum fills happen reliably on beat **4**, rather than the first, second, or third beats of the final bar in a phrase; this offers evidence to the prominence of the fourth beat over the others because when phrasal prominence is realized, it occurs on the final beat of the metrical structure. The prominence of beat **4** in the measure can be represented by adding another level to the metrical structure of (12), as represented in (16):

```
(16)                *       L(2)
        *           *       L(1)
*       *      *    *       L(0)
1   +   2   +  3  + 4   +
```

The grid structure in (16) represents the basic rising rhythmic pattern of the drums and as seen in the bass parts. At the tactus level, the second and fourth beats are the strongest, and between these two beats, beat **4** is the strongest over-all.

The r-s instruments form a clear metrical grid with a weak-strong rhythm. Although there is a desire to mark the beginnings of constituents, which results in the tension between

---

[4] Witmer also defines a fill as a figure “played at points of inactivity or stasis (for example, between phrases, choruses, or solos, or during a sustained note)” (Witmer 2006). However, drum fills, in my data, consistently occur at constituent—metrical or phrasal—boundaries, not in any place with “inactivity of stasis.” Concerning just the drums, therefore, I argue that these fills mark boundaries instead of “fill” as Witmer claims.

actual metrical strength and natural strength, the endings of each constituent is the most stressed in the r-s instruments' role in bop swing. By playing consistently repetitive and regular patterns, the bass and drums delineate a well-structured rhythmic template.

2.2 The harmonic and melodic (h-m) instruments

The rhythmic variation in the non-r-s instruments is much greater since the harmonic and melodic parts do not have regularly periodic rhythmic patterns on the musical surface. While the predetermined roles of the drums and bass are to delineate a near-constant time and succession of beats, the h-m instruments, including the voice, have greater freedom in rhythmic choice without being constrained to certain patterns. Just because the non-r-s instruments do not follow regularly periodic cycles like the bass and drums does not mean that these instruments lack government by any metrical structure and have complete rhythmic freedom. Instead, patterns in the harmonic and melodic parts show a well-defined metrical template that controls the rhythmic freedoms of these instruments.

Based on an analysis of which beats are more prominently and purposely accentuated by the h-m instruments, we can determine the prominence of certain beats over others. For example, in "Goody-Goody" from Fitzgerald's late 1950s record, *Get Happy!* (1957-59), the distribution of accents across beats in Fitzgerald's line is tallied in (17):

(17)

Beat	**1**	**1+**	**2**	**2+**	**3**	**3+**	**4**	**4+**
Accents	1	4	9	4	0	6	3	15

From the chart in (17), Fitzgerald evidently never stresses beat **3**. This lack of stress on the third beat is significant when compared to the greatest number of accents occurring on one beat:

fifteen stresses on **4+.** Like beat **3**, **1** hosts a low number of accents, with only one stress throughout the entire piece. Beats **1** and **3** are the least accented in "Goody-Goody." Thus far, these accentuations provide evidence for a metrical grid congruous to that in (16).

Looking at beat **4+** in (17) complicates the hitherto developed grid. This beat by far carries the most accenting; therefore, it should be considered the most prominent beat of the measure as **3** is the least prominent beat. But, before this, I have not associated stress with any of the off-beats. Incorporating the evidence of **4+**'s prominence could generate either grid represented below (18):

(18) (a)

```
                              *   L(2)
        *               *         L(1)
*       *       *       *         L(0)
*   *   *   *   *   *   *   *     L(-1)
1   +   2   +   3   +   4   +
```

(b)

```
                              *   L(2)
        *                     *   L(1)
*       *       *             *   L(0)
*   *   *   *   *   *   *   *     L(-1)
1   +   2   +   3   +   4   +
```

Problems arise with both representations of the prominence of **4+** above. Since we have no evidence of change in the prominence relationship of the other beats, in (18a), the grid retains its original form from (16), except for the strength of **4+** at the highest level, L(2). This representation, however, violates the "Continuous Column Constraint," (Hayes 34-37), and Jackendoff and Lerdahl's Metrical Well-Formedness Rule 2, which both say that, universally, a beat at any given level must correspond to beats on all smaller levels (Jackendoff and Lerdahl, 69), as illustrated in (19):

```
(19)   *
       *               *
       *       *       *       *
       *   *   *   *   *   *   *   *
```

In (18b), the columns are preserved; however, each beat is not equally spaced, another violation of a *GTTM* rule, MWFR 4 that states that beats should be equally spaced. Although this is not a universal claim from Jackendoff and Lerdahl, it seems unintuitional and insufficient to be resigned to unequally spaced beats: instead, there must be a more appropriate option.

Further complicating this case is the situation of note attacks. Jazz is notorious as a "laid-back" music, characterized by slightly extreme delays or anticipations in the attack of notes so that the beginning point of a note comes either before or after the actual beat on which the note should start. These delays and anticipations do not have the purpose of drawing a listener's attention. They are meant to develop a hazy, obscuring sound and to dull attack points. In "Goody-Goody" and the other pieces, anticipations and delays in attacking a note occur with great frequency on the whole-numbered beats but rarely ever on the off-beats. Not obscuring the onsets of the off-beats almost consistently provides evidence for a different grid from (16), developed by evidence from the r-s instruments. Since here, strength of the whole-numbered beats is attenuated, the off-beats must be the stronger beats—the ones with more stress, and this is represented by the metrical grid model in (20):

```
(20)               *               *  L(1)
           *       *       *       *  L(0)
       *   *   *   *   *   *   *   *  L(-1)
       1   +   2   +   3   +   4   +
```

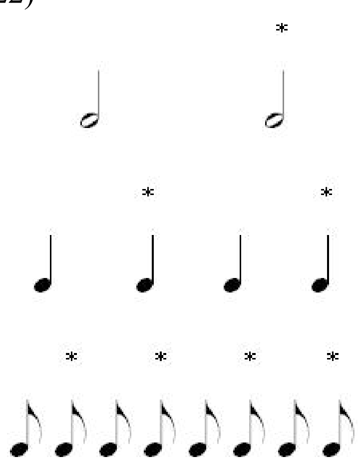
This solves the problem of **4+** being the most prominent; however, the level above the tactus, L(1), in this model does not seem correctly descriptive of jazz. In tallies of anticipations and delays, **2** and **4** consistently are played more on the beat than **1** and **3**. The piece, "Skylark," in particular, has no anticipation or delay on the beat **4** while beats **1** and **3** both have four delays. This would conclude that beats **2** and **4** are more prominent and less obscured than the **1** and **3**, but the model in (20) does not exhibit this relationship. The model in (21) is no solution, either:

```
(21)            *               *       L(1)
            *       *       *       *   L(0)
        *   *   *   *   *   *   *   *   L(-1)
        1   +   2   +   3   +   4   +
```

This model (21) again violates the column constraint discussed above, bringing the grid back to the same problem of representation.

The problem of the structural metrical representation of the h-m instruments' rhythm stems from the evidence of different beats at different levels carrying stress, rather than a cleaner correlation that was evident in the r-s instruments. But, the h-m instruments' rhythm does not completely lack pattern. Like the rhythm of the r-s instruments, the rhythm of the h-m instruments forms a rising pattern, where at L(0) the second eighth note is stronger than the first, at L(1) the second quarter note is stronger than the first, and so on, shown in (22):

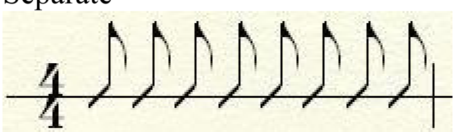
(22)



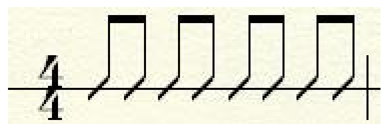
Additionally, remember that beat **4+** is, given accentuation evidence, the strongest in the measure, so the typological characteristics of culminativity and hierarchy—and with these, constituency—in linguistic meter must also be in play here. Beat **4+** may be thought of as a constituent of the entire beat **4**, which is a constituent of the second half of the measure, which is a constituent of the whole measure. In other words, **4+** can be grouped with **4** to form a constituent of a higher metrical level. This idea of constituency in music can sometimes be seen in the actual short-hand notation of music, where eighth notes more often occur barred together than separately (23):

(23) cf.,

Separate

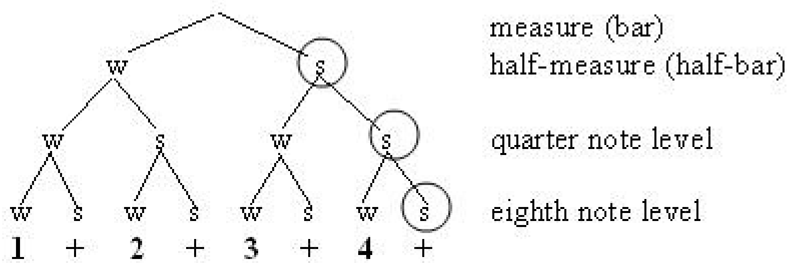


Barred



It is possible to argue, based on culminativity, that beat **4+** is a strong constituent of another strong constituent of another strong constituent; therefore, by being dominated by all strong positions in the measure, beat **4+** is representationally the strongest (24):

(24)

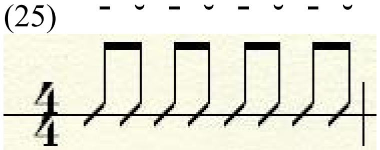


Using a tree diagram solves the representational problem of different beats at different levels carrying stress that the grid was not able to model. Jackendoff and Lerdahl's method of representing musical rhythm fails to take into account the metrical characteristic of constituency, despite the fact that their models incorporate other metrical characteristics, including rhythmicity and hierarchy (Hayes 1995). Prince (1989) argues for the use of trees over grids in poetics for the same reason, that a grid is a "perspicuous representation of purely 'metric' properties" and that a tree represents "abstract… phrasing, with relations of strength circumscribed by constituency" (Prince 48). The h-m instruments of bop swing demonstrate that musical rhythm can be characterized much in the same way that linguistic rhythm can.

I should note here that my proposal for the use of a tree structure representation of musical rhythm in no way conflicts Jackendoff and Lerdahl's reservation of the tree structure for phrasal constituents: the time-span reduction tree. The metrical tree that I posit in (24) represents the phonological property of rhythm in music while Jackendoff and Lerdahl's time-span reduction models the syntactic properties of higher-level phrasal constituents such as harmony in music. These two tree models can definitely coexist without interference: cursory evidence, such as the demarcation of eight or twelve bar phrases in the drums, shows that a time-span reduction tree may be used to model phrasal levels in jazz while a metrical tree is used to model rhythm in jazz.

Based on evidence in accentuations and delaying and anticipating notes, the rhythm of the h-m instruments can be represented as in (24), very much analogous to a representation of iambic tetrameter in poetic meter. However, two other rhythmic patterns in the h-m instruments have not yet been considered: the long-short distinction between eighth notes and triplet rhythm. The former pattern is definitive to swing music, and almost any bop swing piece will have uneven eighth notes. In the long-short pattern, the first eighth note has a longer duration and the second eighth note of the pair has a shorter duration (25)[5]:

(25) ¯ ˘ ¯ ˘ ¯ ˘ ¯ ˘



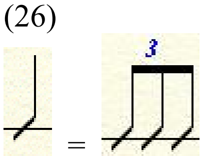
1 + 2 + 3 + 4 +

[5] I have chosen to use the notation of linguistic length [¯, ˘] instead of musical notation of *tenuto* and *staccato* [- , .] because *tenuto* and *staccato* have no effect on the actual duration of the note but rather on how the note is articulated. Linguistic length better represents this long-short length distinction.

This long-short rhythmic pattern causes some problem for the metrical template of h-m instruments' swing rhythm hitherto developed. Having assumed Jackendoff and Lerdahl's MWFR 4 to be true above, the long-short distinction between eighth notes violates this rule. Is it, then, that the unevenness in duration of the eighth notes is a surface phenomenon and should not be represented as a part of the underlying metrical structure? This explanation is unsatisfactory given the omnipresence of the long-short eighth notes in all of the bop swing pieces and the lack of any conditioning environment that causes unevenness to occur. It must be assumed, therefore, that long-short unevenness is an underlying property of bop swing, again resulting in the problem of metrical representation.

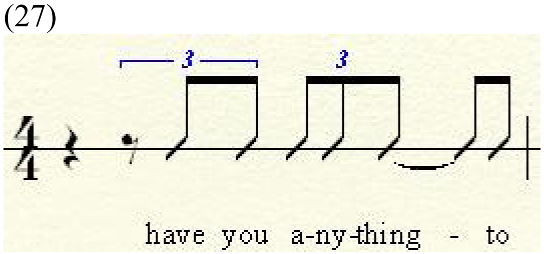
The latter phenomenon of triplets may help in the representational problem of uneven eighth notes. Triplets occur when, instead of splitting a quarter note into two subdivisions, a quarter note is split into three subdivisions (26):

(26)



This ternary splitting occurs more often in the slower bop swing pieces, as the slower tempo probably allows for greater subdivisions and distinctions, since it allots more time to each tactus beat. An example of triplets can be heard in the second bar of "Skylark," transcribed below (27):

(27)

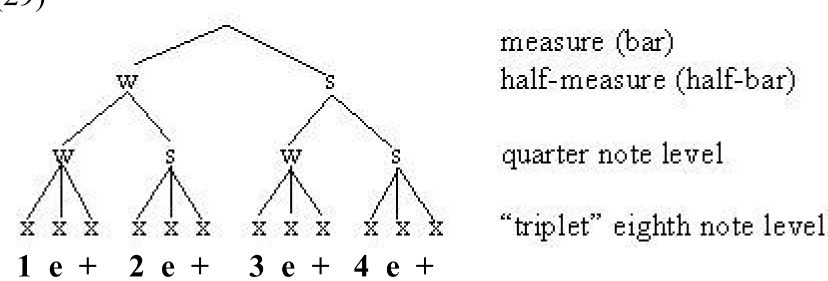


These triplets consistently occur throughout the slower tunes and at some points in faster songs, thus evidence that they, too, are part of the underlying rhythmic structure of bop swing, just as the uneven eighth notes are argued to be. The triplets furthermore raise a representational problem in modeling the rhythm of jazz bop swing and music. Heretofore, the assumed structure of musical meter has generally been a binary-splitting one, as demonstrated by Jackendoff and Lerdahl for Western classical music and as shown in (24) for swing. In order to accommodate underlying triplets, a ternary structure must be possible for music, where feet may contain three lower level positions rather than two in the binary structures posited until now (28):

(28) Binary structure v. Ternary structure

There is no evidence, furthermore, for any ternary splitting on any level above the one the tactus; for example, there are little to no instances where a full measure subdivides into three. Only the lowest level in the metrical model must be modified to accommodate three branches from each quarter note, as revised in (29):

(29)[6]



As for the internal structure of the ternary level, evidence from the frequency of the inception of pitch events, an indicator of prominence noted in *GTTM*, demonstrates second note of each triplet set is the weakest. For example, the chart in (30) shows the frequency of inception of pitch events in each position in "Skylark:"

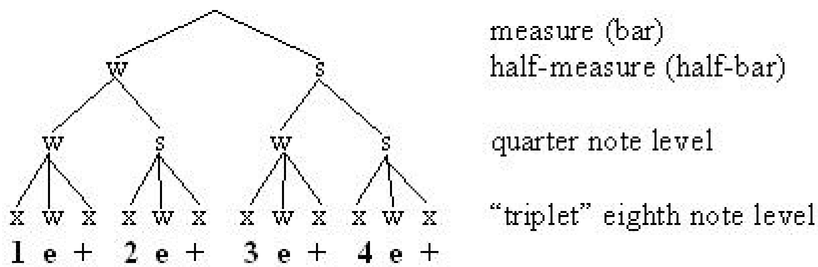
(30)

"Counts"	**1**	**e**	**+**	**2**	**e**	**+**	**3**	**e**	**+**	**4**	**e**	**+**
# of occurrences	22	0	4	24	14	26	32	3	19	10	10	19

[6] The middle note of each triplet will be referred to as "e."

Even though the difference between the frequency of pitch events on whole numbered counts and + counts is not significant, the difference between the frequency of pitch event inceptions on the second note in each triplet and other positions is drastic. Where as many as thirty-two inceptions may occur on a whole-numbered beat, **e** counts consistently hold fewer inceptions, with as few as zero, and at most, ten. This evidence, which patterns consistently in this way in triplet occurrences, generates a model, shown in (31), where the second note of the triplet is undoubtedly weak:

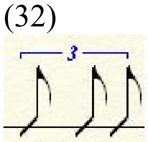
(31)



The frequency in pitch event inception does not provide solid evidence as to the relative prominence of the first or third triplet note position, though it is clear that they are both stronger than the second note in each triplet. To determine the relative strengths of these positions, we must return to the issue of uneven eighth notes.

In bop swing, a pair of uneven eighth notes will have a long-short rhythm, often represented in elementary jazz instruction as two triplet eighth notes tied together and coupled with one triplet eighth note:

(32)



Hence, the first "eighth" note will be longer than the second since underlyingly, the note consists of two triplet eights while the latter "eighth" note only consists of one triplet eighth. In actuality, the surface representation of swing is not that each quarter—tactus level—note is subdivided cleanly into triplets, but this explanation is one used canonically to teach swing rhythm, which demonstrates that musicians conceive of swing in triplets, or a subdivision near to triplets. The long-short distinction concurs with the analysis above that the second triplet eighth in each triplet set is the weakest, because when "eighth" notes are played—that is, when a quarter note is divided into two uneven halves—the second triplet note is the one held silent. Additionally, the first two triplet eighths are often paired together in the long-short distinction while the final triplet eighth stands alone, and the opposite rarely occurs, which is a phenomena that will come into play in deciding the relative prominence of each eighth note.

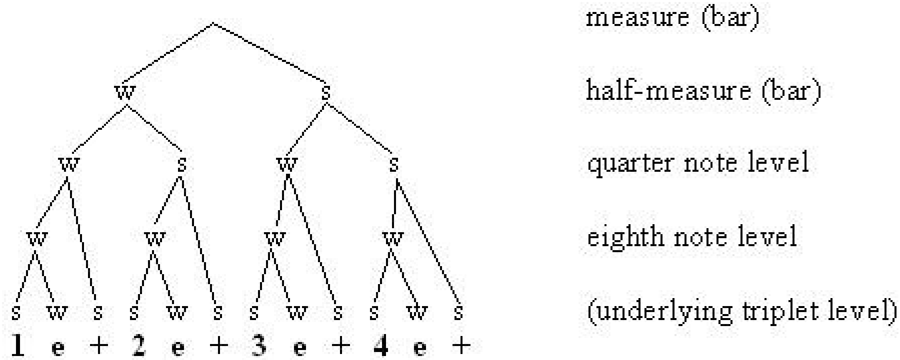
Patterns of accentuation, anticipation, and delay reveal the relative prominence between the long and short eighth notes at the lowest level. Like the argument that led to (24), obscuration of the whole-numbered positions in the anticipation or delay of attack on these notes shows that these positions are weaker than their **+** counterparts. Moreover, the argument for the strength of **4+** from accenting again bears evidence that **+** beats are the strongest overall. The relative prominence among the triplet level, where the first triplet eighth note is the second strongest, the second triplet eighth is the weakest, and the third triplet eighth is the strongest, is represented below:

```
(33)       s     s     s     s     strongest
        s     s     s     s        medial strong
         w     w     w     w       weakest
        1 e +  2 e +  3 e +  4 e +
```

As seen in (33), there is a hierarchical relationship between each of the triplet notes in that the third triplet eighth is stronger than the first and must be represented as a weaker strong node. Furthermore, that the first two triplet eighths are frequently held together to form swung eighth notes but rarely does the reverse occur demonstrates that the first two triplet eighth notes should in some way be paired together in the metrical representation.

In order to accommodate this particular subdivided eighth structure in the metrical model of swing rhythm, taking into account both the ternary splitting of the triplets, their common binary behavior, and their relative prominences, it is necessary to extend our model one "level" further, as shown in (34), to form a metrical structure much like that of anapestic meter in poetry:

(34)



The structure in (34) integrates both binary and ternary subdivision of the quarter note level. At the lowest level—the underlying triplet level—the quarter note subdivides to three metrical positions, with a structure of strong, weak, strong. Moving one level higher, the eighth note level only shows the binary subdivision of quarter note, which correlates with the long-short swung eighth rhythm because its weaker constituent consists of two lower positions rather than one, as the stronger eighth position does. Unlike many metrical structures in poetics that do not necessarily reflect any durational timing, the structure posed here represents actual timing units, for which the two positions under the weak position of the eighth note level actually encode that the first eighth note is longer than the second. Additionally, this model incorporates the metrical distinction of strength between the two stronger triplet eighths by subordinating the weaker strong under a weak position coordinated with the stronger.

2.3 Two models of bop swing rhythm

Thus far, two models of jazz bop swing have been developed: the r-s (rhythm section) grid-based model (16) and the h-m (harmonic, melodic instruments) tree-based model (34). Unlike Western classical music that has only one metrical representation of its rhythm, the two structures of bop swing work in tandem to form the polyphonic rhythmic texture characteristic of jazz and African musical traditions. It is possible, however, to collapse the two models of bop swing into a generalized one, if not for their similarities then at least for the ease of exploring text-setting.

Although there is no argument for constituency structure in the r-s grid model, the structure is similar to the upper levels of the h-m model. Both models illustrate the strength of beats **2** and **4**, with beat **4** being the strongest overall. Only at the lower levels are these two models dissimilar, mainly because the r-s model does not extend below the quarter note level (L(0)). It is not impossible to posit that the r-s instruments play with some constituency in their rhythm; for example, the measure constituent, present in the r-s rhythm, consists in the recurrence of the strongest beat, demonstrating some distinction in constituency.

The only significant difference in the two models that must be reconciled, therefore, is the rhythmic representations at the tactus level and below. For the r-s model, the attack of each quarter note (L(0)) is strong, whereas, for the h-m model, the upbeats are stronger; furthermore, there is the distinction between the inherent strength of the beginnings of higher-level constituents and the metrical prominence of other beats in jazz that is incorporated into the r-s model through the represented strength of the attack of the quarter notes. The strength of the attack of the quarter notes, however, is present at lower levels in the h-m model. At its underlying triplet level, the first triplet eighth is stronger than the second, demonstrating the strength of the quarter note attacks and integrating the possibility for these positions to carry strength in higher level phrases and constituents. The metrical model of jazz bop swing rhythm, therefore, encodes the inherent strength of the higher level beginnings but emphasizes the representation of metrical prominence.

GTTM briefly mentions jazz and dismisses its distinct rhythmic structure as an idiom in which "syncopation [is] a prominent characteristic" (279), defining syncopation as "a situation in which the global demands of metrical well-formedness conflict with and override local preferences" (77). From the jazz bop swing rhythmic model developed here, it is apparent that swing not only involves syncopation as a conflict between inherent strength in higher-level constituent beginnings and metrical prominence but also utilizes a different metrical model from the one that *GTTM* proposes for Western classical rhythm. Thus, the difference between the rhythms of jazz bop swing and Western classical music lies beyond the fact that jazz has a "prominent characteristic" of syncopation—Western classical music and jazz are based on two separate metrical models, one that is a simple grid of metrical positions and the other that encodes more complex features of rhythm. Retrospectively, it is not impossible that Western classical rhythm has constituency structure or even that certain meters in classical music, such as 6/8, involve ternary structure, since these features have now been demonstrated to occur in music and specifically, jazz; however, this possibility warrants further investigation that I will not delve into here.

Now that a metrical model for jazz bop swing rhythm has been developed, later sections will discuss the effects of this rhythmic structure on text-setting.

3. Background: Text-setting

The matching of " 'strong' metrical positions to 'stressed' syllables" is the most basic assumption of the ultimate goal of text-setting, as Halle and Lerdahl (1993) hypothesize. But this hypothesis is by no means simple or basic—nor is it entirely descriptive—of the processes and goals of a text-setting system. The first matter that deviates from a pure stress-to-strong assumption is the important role that non-stress-matching objectives play; for example, the strong preference to have parallel settings in rhyming lines can often override stress matching preferences. Interactions of non-stress-matching preferences and stress-matching ones have been demonstrated to some extent in the work of researchers following Halle and Lerdahl (e.g., Hayes and Kaun 1996, Hayes and MacEachern 1997, Kiparsky 2006, etc).

The second issue taken up here is the practical realization of stress-to-strong goals in jazz bop swing. Previous text-setting research has only dealt with the setting of linguistic text to one musical grid. Given the tandem models for swing rhythm developed in this paper, text-setting in jazz must wrangle with competing positions of strength—inherent, left-boundary marking strength versus metrical strength in the rhythmic grid and tree (see §2.1). Thusly, for an accurate model of text-setting in jazz bop swing, the basic definition of text-setting as a stress-to-strong phenomena is insufficient. The remainder of this paper explores the goals that text-setters consider when pairing text to tune: stress-matching, phrasal alignment, and phrase- and line-external preferences. A system for modeling text-setting intuitions in jazz bop swing follows from the interplay between various competing goals of matching the linguistics to the music.

3.1 The Serial Approach: Halle and Lerdahl (1993)

The first stab at a generative analysis of text-setting, given by Halle and Lerdahl (1993), was not a system of competing goals in matching the linguistics to the music; rather, they attempt to describe the process of text-setting using a rule-based analysis, grounded in stress-to-strength correspondence. Central to their description of the intuitions of a text-setter is the Text-setting Algorithm, which assigns linguistic stress to musically strong positions via left to right correspondence, shown in (35).

(35) The Text-setting Algorithm

Step 1. Assign all accented syllables to available L(0) beats from left to right.

Step 2. Assign syllables to L(-1) beats from left to right.

Step 3. Assign syllables to L(-2) beats from left to right.

(15 Halle and Lerdahl)

In addition to this process, Halle and Lerdahl posit Text-setting Well-formedness Rules (TWFR), modeled after Jackendoff and Lerdahl's Metrical Well-formedness Rules (MWFR). The rules function in concordance with both the Text-setting Algorithm and the MWFRs, and the application of rules and the algorithm seems cyclic: Step 1 first applies and then all of the TWFRs and MWFRs apply; Step 2 then applies, again followed by the well-formedness rules; and so on, so forth.

Halle and Lerdahl's method of modeling the text-setting process in rules and serial rule application works well for deriving some lines but not all. Hayes (2005) points out the main problems with their analysis: leftward greed, squeezing unstressed syllables, and handling more than four stresses and consecutive stressed syllables (Hayes 9-11) . The Text-setting Algorithm, therefore, proves too simplistic, concerned only with left-to-right stress association. Even though this linear correspondence seems to emulate the way a person reads text and music from the left to right, it causes several of the problems that Hayes (2005) notes. With fewer than four stressed syllables in a linguistic phrase (henceforth, l-phrase), leftward greed occurs instead of spreading the l-phrase out through the entire musical phrase (m-phrase) allots, the l-phrase attracts heavily to the left side.

(36) Leftward greed, adapted from Hayes (2005)

```
            *               *               *               *
    *       *       *       *       *       *       *       *
    *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
*   I     pró-mised her I'd már-            ry              her.
```

Left to right association in the Text-setting Algorithm also causes the squeezing of stress-less syllables, similar to leftward greed, with the stressed syllables mapped already from left to right—and therefore immovable—any extra stress-less syllables must be mapped in between, which may result in syllable squeezing:

(37) Squeezing stress-less syllables, adapted from Hayes (2005)

```
            *               *               *               *
    *       *       *       *       *       *       *       *   *
    *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
    * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
*   En-     quir-   ingforhisa-lá- dy  ó
```

Another cause of the main problems with Halle and Lerdahl's (1993) text-setting analysis is its overly deliberate interpretation of musical strength. That is, the requirement that the linguistically strongest syllable must be mapped to the musically strongest beat limits the flexibility of the algorithm to handle four or more stressed syllables in a l-phrase or consecutively stressed syllables. When it comes to too many stressed syllables in the l-phrase, the Text-setting Algorithm runs out of musically strongest beats in which to map the syllables:

(38) Handling four or more stressed syllables, adapted from Hayes (2005)

```
            *               *               *               *
    *       *       *       *       *       *       *       *
```

```
   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
 * I       pláced  my      báck    a-      gáinst  the     old

           *               *               *               *
   *       *       *       *       *       *       *       *
   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
           gár-    den     gáte.
```

Related to this problem is one of handling consecutive stressed syllables. The deliberate interpretation of strength is once again problematic, because it forces consecutive stressed syllables to spread through the m-phrase and, as a result, compresses the rest of the phrase uncomfortably:

(39) Handling consecutive stressed syllables, adapted from Hayes (2005)

```
           *               *               *               *
   *       *       *       *       *       *       *       *
   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
 * To      córt            yóung           mái-dens I  was bént.
```

Halle and Lerdahl's generative text-setting model under-generates—it cannot handle certain forms of l-phrases. In text-setting reality, l-phrases are not constrained by only having four stresses that are non-consecutive and evenly spaced throughout a given l-phrase.

3.2 Towards an OT-based analysis

The crucial conclusion of experimental, grammaticality-judgment text-setting research is that text-setting is productive, as Hayes and Kaun (1996) found with their "Drunken Sailor" study. A native speaker can set almost any given line of text—provided that it's grammatical English—to a given m-phrase. Thusly, there must be more at work than pure stress-to-strong matching and that Halle and Lerdahl's left-to-right stress association algorithm needs modification.

As a solution to these problems, Hayes (2005) proposes a constraint conflict approach as a way to model the competing goals of the text-setting process, drawing from recent developments in phonological analysis. Constraints allow us to characterize the various goals of a system, which may include stress-to-strength, phrasal alignment, and line-external goals. The conflict between the constraints follow from the priorities of the text-setter: of course, the text-setter would like to satisfy as many goals as possible, but he must make "strategic sacrifices" (11 Hayes 2005). Because of its unique ability to characterize and prioritize goals and preferences, Optimality Theory (OT) makes for an ideal framework in which to deal with text-setting.

An Optimality-theoretic approach to text-setting is just that: an approach. Optimality Theory—as applied to text-setting in Hayes's work—has constraints and constraint rankings that choose between the possible outputs. But, Optimality-theoretic text-setting differs from Optimality Theory in two vital ways. First, Hayes originally meant for OT in text-setting to judge the well-formedness of a given text-to-tune matching (Hayes 2007, voce). That is, at the level that we are concerned with—where text and tune have already been independently generated—OT steps in and judges how well these are matched. Second, any Optimality-theoretic system for text-setting must be more flexible than linguistic OT. Music is, after all, an art form, even if we assume that whatever the composer and performer does falls within a certain grammatical range. Previous text-setting research has been able to push this fact to the background by exploring text-setting in folk song, which has to be, by nature, immediately accessible to many people at once, and so it remains quite regular. Composed music, however, is not as rigid, and jazz bop swing, in particular, hyper-elevates the amount of artistic variation due to its improvisational nature. Again, this is not to say that jazz is completely irregular—it still functions within the limits of what defines jazz; however, any Optimality-theoretic analysis that we use of text-setting in jazz bop swing must be more flexible than that of OT as applied to language.

Given these caveats, OT remains still the best linguistic framework developed thus far in which to begin an analysis of jazz bop swing text-setting. The basic assumptions of OT still resonate. OT-text-setting has constraints that capture the various goals of a system. It has competing constraints that demonstrate prioritized goals. I would even go as far as to speculate that the ranking of constraints, like in language, plays a role in determining different "grammars," which results in different musical systems and traditions. As a part of his analysis of the "Drunken Sailor" song, Hayes (2005) gives constraints and a constraint ranking to model his folk song data. So from this, the interesting question becomes the constraint rankings of jazz bop swing text-setting.

The next section (§4) explores the goals of the text-setting jazz bop swing system and proposes constraints and a ranking of these constraints for jazz bop swing.

4. Text-setting in jazz bop swing

As stated above in §3.2, the input of an Optimality-theoretic text-setting system differs from inputs into traditional OT tableaux in that the lines of text and the lines of tune have already been generated. The input, therefore, includes two components: the l-phrase and the m-phrase. M-phrases are comprised of sets of four-beat musical measures—these are usually powers of two (2^x) in length. The most common m-phrase length is four bars, standard for most medium-fast and fast jazz bop swing tunes. For simplicity, most of the input m-phrases in this study are four-bars long, except that some complications do arise. Additionally, to limit the scope of this paper, I am primarily concerned with the level at which the text to tune matching occurs and where the set line's grammaticality is judged; thus, I will only briefly speculate in §5 on what constrains the l-phrase and m-phrase inputs.

An important feature that will fall out from the necessities of the system I show here is the utilization of gradiently violable constraints, which allow us the flexibility we need for describing this as a regulated art from, as discussed above in §3.2. Gradiently violable constraints, though argued by some to be unnecessary in Optimality Theory (McCarthy 2003), buys us two critical advantages in the case of text-setting. For one, use of gradiently violable constraints will simplify our jazz bop swing text-setting system by cutting down the number of constraints necessary. And then, probably most notably, gradiently violable constraints mirror the relative nature of our main tree-based rhythmic model of jazz bop swing by also allowing relativity in interpretation. We will see this most clearly in the stress-matching category of constraints.

The following sections will exhibit the constraints necessary to satisfy a basic, general description of the text-setting process in jazz bop swing, classified into stress-matching constraints, phrases and phrasal alignment constraints, and line-external constraints. Then, I will discuss the constraints' interactions.

4.1 Stress-matching

The first question to ask and the most basic assumption of text-setting is stress-to-strong matching—how linguistic stress matches up with musical prominence. Even though Halle and Lerdahl's initial assumption that the stressed linguistic syllables correspond to the strongest musical beats proved incorrect, the tendency for stress to correspond to strongest still holds. If this tendency cannot be satisfied, however, the system will go for the second available strongest musical beat in which to map the stressed syllable. So, it suffices to modify the claim: stressed linguistic syllables correspond to the strongest available musical beats.

Hayes's (2005) analysis incorporates this necessary sacrifice in the form of tiered constraints:

(40) from Hayes (2005)

	Strong Position	**Medium Position**	**Weak Position**
Stressed syllable		*STRESSED IN M	*STRESSED IN W
Stressless syllable	*STRESSLESS IN S		*STRESSLESS IN W
Absence of syllable	*NULL IN S	*NULL IN M	

Basically, the constraints from Hayes's chart in (40) come from the generalization that the stressed syllable wants to be in the strongest available position. Because stressed syllables, if given the choice, would go on the strongest beat, we have the constraints *stressed in M and *stressed in W. Then, because the stressed syllables want to go into the next available strongest beats if they cannot associate with the strongest, *stressed in M must crucially outrank *stressed in W.

It is true, too, that for text-setting in jazz bop swing, the stressed syllables associate with the strongest musical beats. The contrast between the frequency of polysyllabic main stress, which is not conditioned by phrasal environment and therefore the most reliable to study, in beats **2**, **4**, **2+**, **4+** and beats **2e**, **4e** demonstrates this tendency:

(41) from "Goody-Goody"

Beats	**2**	**2e**	**2+**	**4**	**4e**	**4+**
Polysyllabic main stress	4	0	4	10	0	4

The number of polysyllabic main stresses that fall on beats **2e** and **4e** is consistently none, whereas more polysyllabic main stresses occur on beats **2**, **2+**, **4**, and **4+**. Similarly, an investigation of stressed syllables, including polysyllabic main stresses, yield the same results:

(42) from "Goody-Goody"

Beats	**2**	**2e**	**2+**	**4**	**4e**	**4+**
Polysyllabic main stress	10	0	8	10	0	14

Again, no stressed syllables associate with beats **2e** and **4e**, whereas beats **2**, **2+**, **4**, and **4+** carry the most stressed syllables. The data does not really distinguish between **2**, **4** and **2+**, **4+** in which stressed syllables favors, and since these beats are only one tier away from each other in the rhythmic tree of jazz bop swing, this demonstrates little difference from the stressed syllable for the strongest musical beats or the second available strongest. So, given this data thus far, text-setting in jazz bop swing shares the same stress-matching goal as text-setting in Western classical music and folk song because both traditions want to set stressed syllables into the strongest available musical beats possible.

The one problem in this analysis is the data for the frequency of stressed syllables in beats **1** and **3**, shown in (43) below.

(43) from "Goody-Goody"

Beats	**1**	**e**	**+**	**2**	**e**	**+**	**3**	**e**	**+**	**4**	**e**	**+**
Stressed syllables	24	0	1	10	0	8	17	0	3	10	0	14

The data in (43) presents the analysis developed thus far with the evidence that beats **1** and **3** attract as much linguistic stress as—if not more than—beats **2**, **2+**, **4**, and **4+**, which are supposedly the strongest beats according to the rhythmic tree. Remember, however, that there was a similar problem in the development of a rhythmic model for jazz bop swing, discussed in §2.1 with the bass pattern that emphasized the inherent initial strength of left-edge boundaries. Due to its poly-rhythmic texture with competing strengths, jazz bop swing has two strong, separate, possible landing sites for stressed syllables: the metrically strong **2+** and **4+** and the inherently strong **1** and **3**. Thusly, the beats that mark left-edge boundaries and that are, because of their function, inherently strong, can be thought of to be on the equal metrical tier as the metrically strongest beats, which then puts them at the top of the list of possible stressed syllable landing sites.

To turn now to the constraints that model these generalizations on stress-to-strength matching, it is evident that the constraints needed for text-setting in jazz bop swing will be similar to those in folk song, since the main goal remains the same. It would therefore be possible to use Hayes's *STRESSED IN M, *STRESSED IN W, et cetera constraints; however, these constraints fail to capture the very relative nature of the text-setting system: stressed syllables go into the strongest beat available, so, if the strongest beat were not available, the syllable would move on to the next strongest and so on. The stressed syllable could even land on the weakest beats. What Hayes's constraints generalize is that stressed syllables do not often fall on medial strong or weak beats, but Hayes's constraints do not tell you that the landing of a stressed syllable onto a weak beat is related to the fact that the strongest beats just were not available. For Hayes, this sort of follows from the constraint ranking, but I want to assert an alternative analysis in utilizing gradiently violable constraints.

A result of a gradiently violable constraint system is a simplification of what Hayes (2005) presents, as pictured in (40) above. Hayes uses three levels of linguistic strength: stressed, stressless, and null. He assigns two star-struc constraints to each level of linguistic stress to denote which musical beat is the preferred landing site of each kind of syllable. This approach works well until the introduction of the "NULL" syllables, where Hayes's analysis runs into a problem of cognitive reality. Even though the constraints *NULL IN S and *NULL IN M correctly derive the optimal outputs for Hayes, in reality, it is impossible to set a null syllable—essentially, a syllable that does not exist—to a beat. Using gradiently violable constraints will help to remedy this problem and, in the process, cut the number of stress-matching constraints in half.

Instead of Hayes's six constraints, a stress-to-strength-matching system based on gradient violability only needs three constraints (44-46):

(44) **POLY-SYLLABIC MAIN STRESS IN M-STRONG** (gradiently violable): Polysyllabic main stress should be in a musically strong position.

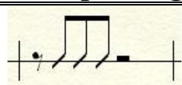
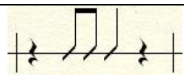
(45) **L-STRESS IN M-STRONG** (gradiently violable): Linguistically stressed syllables should be in musically strong positions.

(46) **L-STRESSLESS IN M-WEAK** (gradiently violable): Linguistically stressless syllables should be in musically weak positions.

These constraints in (44-46) accumulate violations based on the degree of mismatch between the linguistic stress and musical strength. For example, (45) requires that linguistically stressed syllables be mapped into musically strong positions, and, if given the choice, the strongest beat available; consequently, if the stressed syllable falls on a beat that is not the strongest, it receives a violation for each weaker level. Interpreted within the rhythmic tree, this means that for every weak node that the syllable steps down to from the top, it receives a violations. Similarly, since I argue above that inherent strength resides on at least the same tier as metrical strength, the same violation assignment applies to mismatches between linguistic stress and inherent, musical strength.

Consider the setting of the polysyllabic word "marvelous" from "Too Marvelous for Words." Given in the tableau in (47) are three possible output candidates—each musical note represents one syllable.

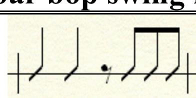
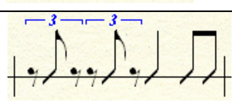
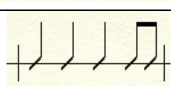
(47)[7]

/marvelous, 1-bar bop swing rhythm/	POLY-σ MAIN STRESS IN M-STRONG
(a) 	*!*
(b) 	*!
☞ (c) 	--

(47a) receives two violations because the main stress of "márvelous" is placed in beat **1+**, which, according to the rhythmic tree, is two down-steps to weak nodes. (47b) incurs one violation because the main stress of the word occurs in beat **2**, one weak node down from the strongest beat. The optimal candidate, (47c) places the main stress in beat **1**, which is the strongest position of inherent strength—equal to the strongest metrical position—and thusly wins.

To demonstrate the other two gradiently violable stress-matching constraints, consider the l-phrase "love song from the birds," which carries stress on "love," "song," and "birds." The tableau in (48) below gives a few possible candidates and unranked constraints:

(48)

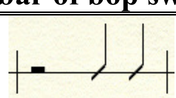
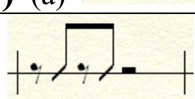
/love song from the birds, 1-bar bop swing rhythm/	L-STRESS IN M-STRONG	L-STRESSLESS IN M-WEAK
(a) 		***, ***!
(b) 	*!***, ***	**, ***
☞ (c) 		**, ***

[7] For simplicity, I have only allotted one bar in the input m-phrase for some of these examples. For discussion of how the length of the input m-phrases are determined, see §4 and 5.

The candidate (48b) accumulates seven violations with the constraint **L-STRESS IN M-STRONG**: four for setting "love," a stressed syllable, into beat **1e**, and three for setting "song," a stressed syllable, into beat **2e**. (48b) dies with only one violation since the other candidates given here do not violate **L-STRESS IN M-STRONG**. As for (48a), the other non-optimal candidate, it violates **L-STRESSLESS IN M-WEAK** six times: three times for placing "from," an unstressed syllable, in beat **3+**, which has three strong nodes, and three times again for placing "the," another unstressed syllable, in beat **4**, which has three strong nodes above it. (48a) dies on its sixth violation, since it is tied with (48c) up until that point. (c) accumulates five violates for placing "from" on beat **3** and "the" on beat **4**. Even if the two constraints were ordered oppositely, (48c) would still win because (48b) also collects five violations for **L-STRESSLESS IN M-WEAK**.

With these three stress-to-strength constraints, it is possible to rule out many but not all of the non-optimal text-to-tune candidates. Consider the tableau in (49):

(49) from "That Ol' Black Magic"

/in its, 1 bar of bop swing/	**L-STRESSLESS IN M-WEAK**
(☞) (a) 	**, ***
(b) 	**, ***

Both candidates, though (a) is supposed to win, tie in violations; moreover, the other stress-matching constraints do not apply here, so there must be goals beyond pure stress-to-strength matching.

4.2 Phrases and phrase matching

Phrasal constraints work in tandem with stress-matching ones to select optimal text-to-tune candidates. Some phrasal constraints, we will see, crucially outrank stress-matching constraints, which means that stress-to-strength correspondence is not the only goal a text-setting system is concerned with. Phrasal constraints help to regulate the distribution of linguistic events—syllables—throughout a time span—the m-phrase; they also determine where a l-phrase begins and ends in relationship to the m-phrase. Phrase-related constraints are split into a few major categories of phrasal constraints that govern specific components of each phrase: left edge, right edge, alignment, and internal constraints.

4.2.1 Left edge phrasal constraints

In jazz bop swing text-setting, anacrusis is encouraged, and, in almost every line, there is at least a syllable that occurs before the first measure of the four-bar phrase. This is known in music as the pick-up note or notes, and most text-setting researchers have hitherto treated these syllables as part of the actual line. Doing so, however, cripples the analysis by conflating the music and the language. The pick-up syllables are part of the l-phrase and must be analyzed so, since something like “so, you” cannot exist on its own as a grammatical l-phrase; conversely, the pick-up syllables are not part of the m-phrase since, underlyingly, text has not yet been set to tune. Treating pick-up syllables as inherently part of the m-phrase, therefore, is troublesome because they do not actually exist as part of the metrical, musical input so they should not be analyzed as such. These pick-up syllables should be treated as initial, “extra-metrical” syllables which the jazz bop swing text-setting system encourages, and the actual phrase begins on beat one of the first measure of the m-phrase. I am not sure what the motivation behind preferring anacrusis is, but, to only speculate, would suspect that it allows more metrical positions and stretches out the line due to limitations put on the strictness of right edges (see §4.2.2)

Because there is such a drive to have anacrusis at the left edge of lines, the system must have a constraint as in (50) below:

(50) **HAVE PICK-UP**: have an incomplete measure before the first measure of each m-phrase.

But, we don't often find that pick-ups will consist of more than one or two syllables, which motivates the constraint in (51).

(51) ***COMPLEX PICK-UP** (gradiently violable): do not have more than one note in the pick-up.

Additionally, not everything may go in pick-up syllables—any stressed syllables, whether or not they are poly-syllabic main stress—are rarely found in the pick-up; thusly, the constraint in (52).

(52) ***STRESS IN PICK-UP**: no stressed syllables in anacrustic measures.

In theory, all of these constraints can be gradiently violable, much like the stress-matching constraints; yet, it is unnecessary to use **HAVE PICK-UP** and ***STRESS IN PICK-UP** as gradiently violable constraints. Not having a pick-up at all will violate **HAVE PICK-UP**, so we don't need gradient violability because we don't care how much of a pick-up there isn't. Similarly, since ***STRESS IN PICK-UP** broadly covers all stressed syllables—no matter how strong the stress is—there is no need for gradient violability in this constraint. Gradience is necessary, however, for the constraint ***COMPLEX PICK-UP** since it is possible to have bigger and bigger pick-ups: if an anacrustic measure consists of two syllables, then it violates ***COMPLEX PICK-UP** once, if it consists of three syllables, it violates ***COMPLEX PICK-UP** twice, and so on.

To demonstrate how these constraints work, take the hypothetical input in (53). The linguistic entry to the input is three syllables, with stress on the third syllable, as in "from the heart." **HAVE PICK-UP** must be ranked high, since there is almost always at least one pick-up syllable.

(53)[8]

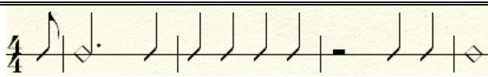
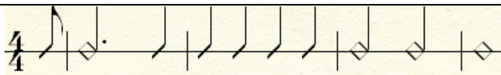
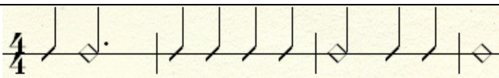
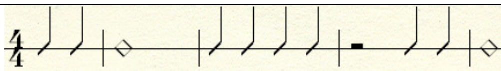
/σ σ σ́, 1 musical bar + pick-up measure/	HAVE PICK-UP	L-STRESS IN M-STRONG	*STRESS IN PICK-UP	*COMPLEX PICK-UP
(a) **σ | σ σ́**		*!*		
☞ (b) **σ σ | σ́**				*
(c) **σ σ σ́ |**		*!	*	**
(d) **| σ σ σ́**	*!	**		

Candidate (d) fatally violates the highest ranked constraint, **HAVE PICK-UP**, since there is no anacrusis. Opposite of this is candidate (c), which has three pick-up syllables, pushing everything to the front. Ignoring the stress-matching constraint in this tableau, (53c) would violate ***COMPLEX PICK-UP** and ***STRESS IN PICK-UP**; this candidate violates ***COMPLEX PICK-UP** twice, since there are two pick-up syllables in addition to the allowable one. The difference between the remaining two candidates, (53a) and (53b), lies in the fact that candidate (a) allows a stressed syllable to land on a beat that is not the strongest, thereby violating the constraint **L-STRESS IN M-STRONG**. This demonstrates the interaction between stress-matching constraints and other constraints, since, without stress-matching, (53a) would win, as (53b) violates ***COMPLEX PICK-UP**. Looking at the data, one finds an overwhelming preference to place the first stress of a l-phrase into the first, non-anacrustic beat of the m-phrase, something that is modeled by the ranking of **L-STRESS IN M-STRONG** over ***STRESS IN PICK-UP** and ***COMPLEX PICK-UP**. The winning candidate is, therefore, (53b).

As another example of the interaction between stress-matching constraints, consider the tableau in (54): the l-phrase input is /That ol' black magic's got me in its spell/ and the m-phrase input is a four bar rhythmic, jazz bop swing phrase.

[8] Each syllable represents one beat.

(54) from “That Ol’ Black Magic”

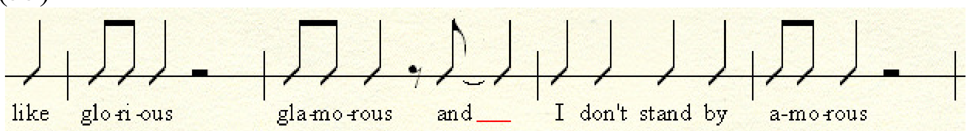
/That ol’ black magic’s got me in its spell, 4 bar m-phrase/	*STRESS PRE-M-HEAD	POLY-σ MAIN STRESS IN M-STRONG	L-STRESSLESS IN M-WEAK
☞ (a) 			**** (that), **(in), *** (its)
(b) 			**** (that), **** (in),**! (its)
(c) 		*!* (má-)	
(d) 	*! (ol’)		

***STRESS PRE-M-HEAD** rules out candidate (54d), since “ol’,” which is stressed to some degree because it is a content word, falls on beat **4** before the anacrustic bar. Candidate (54c) violates **POLY-σ MAIN STRESS IN M-STRONG** twice, with the main stress of “mágic” falling on beat **2**, subsumed under two weak nodes. Then, (54a) and (54b) nearly tie in , but (b) violates the constraint one more time than (a) does, with “in,” a stressless syllable mapped into beat **1**, which is, due to inherent strength properties, one of the two strongest positions in the music. Thus, “in” in (__b) incurs four violations.

4.2.2 Right edge constraints

While the initial, anacrustic measures of lines in the text-setting system of jazz bop swing are discouraged to carry stress, right edges attract stress. The motivation for stricter stress-matching at the end of lines might have to do with finality—end on a strong syllable and beat lends finality to the line and marks the right edge. The robust pattern, therefore, is to end the line by matching the final linguistic stress with beat **1** of the final measure of a m-phrase or beat **4+** of the previous measure, demonstrated by an example from “Too Marvelous for Words” in (55).

(55)



The reason that beat **4+** of the final measure is never preferred as a landing site for this final stress probably lies in the fact that the positions at the end of the final measure also act as the positions for anacrusis in the following m-phrase. And since the Have Pick-up constraint ranks high, the system falls back to the next available strong positions from the end: beat **1** of the final bar and beat **4+** of the preceding one. We also know from the discussion above that beats **1** and **4+** share the distinction of being the strongest beats of the measure, inherently and metrically.

The constraints that are necessary to model this finality are laid out in (56 - 58).

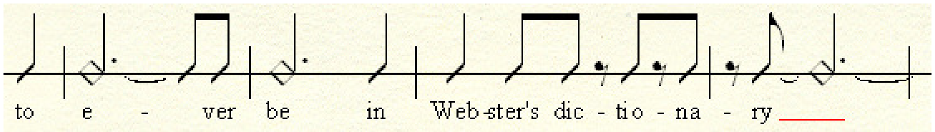
(56) **PHRASE FINAL POLYSYLLABIC STRESS IN M-STRONG**: Have phrase-final polysyllabic stress in a musically strong position.

(57) **PHRASE FINAL STRESS IN M-STRONG**: Phrase final stress should be in a musically strong position.

(58) ***EXTRAMETRICALITY**: No extrametrical syllables.

Given the elevated status of polysyllabic stress discussed above—that it does not depend on phrasal stress—the matching of right-most polysyllabic stress tends to be stricter than the matching of right-most stress, regardless of word syllabicity. The constraint **PHRASE FINAL POLYSYLLABIC STRESS IN M-STRONG** captures this generalization, while a more general **PHRASE FINAL STRESS IN M-STRONG** allows for the more general tendency of finality. In fact, many polysyllabic main stresses can be found in beats **1** and **4+**, but this is not true across all lines; take, for example, the line in (59).

(59) from “Too Marvelous for Words”



So, **PHRASE FINAL POLYSYLLABIC STRESS IN M-STRONG** is a better and more general constraint. As a counterpart to these phrase-final constraints, I have added ***EXTRAMETRICALITY**, since extrametricality after the final beat is often found in order to satisfy higher ranking Phrase final stress in m-strong constraints; because initial stress is characteristic of English, the output may have extrametrical syllables trailing after the main stress, which is set on beats **1** or **4+** for finality.

4.2.3 Phrasal alignment

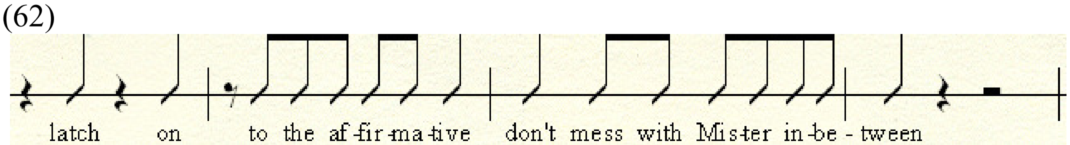
Going along with left and right edge constraints are alignment constraints that govern the alignment of the l-phrase to the m-phrase:

(60) **ALIGN L (l-phrase, m-phrase)**: the linguistic phrase should be no later than beat **1** of the first measure of a m-phrase.

(61) **ALIGN R (l-phrase, m-phrase)**: the linguistic phrase should not run over the allotted musical phrase length or run into the beginning of the next phrase.

In terms of left-edge alignment, it is rare to find any l-phrase that begins later than beat **1** of the first measure of a m-phrase, mostly because **HAVE PICK-UP** is such a strong and high-ranking constraint. The few exceptions that we find to this constraint are artistic ones. In “Accentuate the Positive,” several lines start after beat **1**, usually no later than beat **2**:

(62)



This is allowed because of the Verb Phrase ellipsis occurring throughout the song, where "You got to" is said only once at the beginning of a stanza and all of the subsequent lines correspond to it. It's as if at the same time there is VP ellipsis, there is musical ellipsis.

The right edge alignment constraint in (61) is akin to Bruce Hayes's ***RUN ON** constraint, except that alignment constraints do the same work as ***RUN ON**, so I have chosen to use **ALIGN R (l-phrase, m-phrase)** here.

4.2.4 Phrase-internal constraints

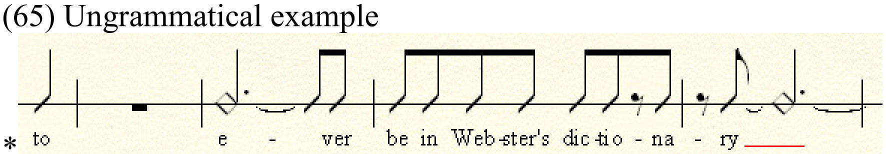
The constraints that govern the edges of a text-setting line, in cooperation with stress-matching constraints, nearly suffice to describe many of the lines, since their demands on both the left and right edges of lines and the stress-to-strength matching already control and determine much of what happens in the middle. There are only two more phrase-internal constraints necessary: ***LAPSE** and **MAJOR SYNTACTIC BREAK @ EDGE**, both defined below.

(63) ***LAPSE** (adapted from Hayes 2006): Have at least one syllable per m-phrase measure.

(64) **MAJOR SYNTACTIC BREAK @ EDGE**: Line edges may only be major syntactic breaks (and these usually correspond with major phonological breaks).

The first constraint in (59), ***LAPSE**, prevents the formation of ungrammatical lines with significant empty spaces line-internally and also between lines, as in the ungrammatical example below:

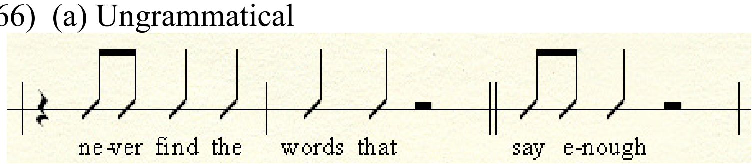
(65) Ungrammatical example

*


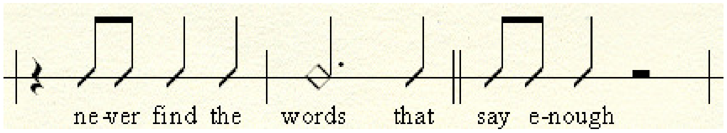
***LAPSE** insures the avoidance of squeezing syllables together, since long lapses would mean that the text-setter would have to fit the l-phrase into less available m-phrase space. The tricky thing about ***LAPSE**, though, is that the satisfaction of this constraint is not always completed by the individual melody. Often, when it seems as if the voice or melody has violated the constraint, other instruments step in and play melodically in the space so that the music will never be empty. When jazz musicians accompany other players, they play small improvisations in spaces where the melody has a lengthy pause; this improvisation is aptly called a "fill," to make up for otherwise empty and stagnant space in the music. Musicians are usually instructed to fill when learning to play jazz.

The second constraint in this set is not as much about where the l-phrase can go in relationship to the m-phrase in text-to-tune setting but rather, what can go into the l-phrase in the first place. I will return to this topic in more detail in §5, but the pattern of major syntactic breaks corresponding with musical line breaks is too robust to leave out of the constraint grammar here. As stated in the definition of this constraint in (64), major syntactic breaks, which usually correspond with major phonological breaks, match up to the musical line breaks; consequently, it is possible to find sentences, beginnings of verb phrases, or complement phrases at line breaks. It is rare, however, to find musical line breaks that match to minor syntactic breaks or no syntactic breaks at all; for example, it is common to find a complete complement phrase as a stand-alone l-phrase but very rarely do we find a complement phrase and its complementizer separated by a musical line break, as demonstrated in (66):

(66) (a) Ungrammatical

*


(b) Grammatical



In the ungrammatical form of (66a), the complementizer "that" comes with the preceding musical line, whereas with the grammatical (66b), "that" is a pick-up that belongs to its complement phrase.

4.3 Line-external constraints

Previous text-setting researchers have given the most time to discussing and analyzing line-external generalizations, with Hayes and MacEachern's (1997) line types, stanza structures, and saliency and parallelism effects. Hayes and MacEachern's (1997) line types are based on different line edings and rhythmic cadences that they found in their folk song corpus. These lines then, in combination, form stanzas, and, depending on the patterning of the lines, produce a specific number of different types of possible couplets and stanzas. These line-external constituents of the couplet and stanza—made of two and four lines, respectively—are governed by saliency and parallelism issues that pare down the number of possible couplet and stanza structures to the ones that actually occur in the data.

Well-classified line types seem to be a phenomenon especially exclusive to folk song verse, since I have found no evidence thus far for specific line types based on rhythmic cadences—the behavior of the final syllables in a line—found in jazz bop swing. I suspect that this is due in part to the artistic nature of jazz, which, as discussed previously in this paper, allows for wider and less-regularized behavior. This is not to say that saliency and parallelism play no role in jazz—not in the slightest; saliency and parallelism actually play a significant part in the final output of a text-to-tune setting, only the clean classification of specific line, couplet, and stanza types is less possible.

Saliency, as defined by Hayes and MacEachern (1997) is either when the end of the last line in a stanza is shorter than the other lines the stanza or when all of the non-final lines have the same cadence and line type. Hayes and MacEachern put it this way:

(67) A metrical constituent is salient if

(a) it's final rhythmic cadence is more cadential than all of its non-final cadences.

(b) all of its non-final cadences are uniform.

18[9] Hayes and MacEachern 1997

In this paper, I will not develop a full view of saliency effects in jazz bop swing text-setting, since I was instead primarily concerned with the basic phrase- and metrics-related goals of the text-setting system. I will remark, however, that it does seem as if saliency plays some part—even if a small part—in the formation of couplets and stanzas. Take, for example, the formation of a stanza in "Something's Gotta Give," in (68) below:

(68) from "Something's Gotta Give"

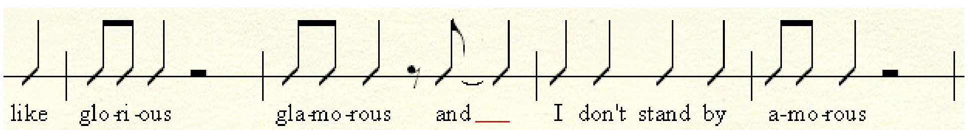
When an irresistible force such as you
Meets an old immovable object like me
You can bet as sure as you live
Something's gotta give, something's gotta give, something's gotta give.

Based purely on syllable count, the pattern for this stanza is 11, 11, 8, 15. Even though this does not follow Hayes and MacEachern's (1997) definition of saliency since the final line is not shorter than the others, there is still a sense of saliency coming from the contrast between eight and fifteen syllables in the second couplet. How saliency works in jazz bop swing and non-folk song traditions warrants further future investigation.

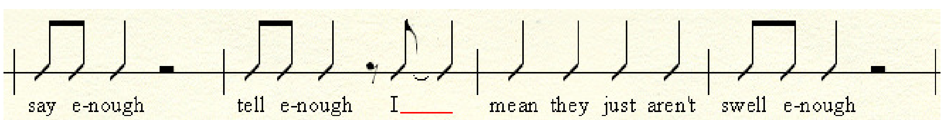
[9] I am using the pre-print version of Hayes and MacEachern, available from Hayes's website.

Similarly with parallelism, Hayes and MacEachern's original formulation of the constraint does not apply very well to jazz bop swing due to the dearth of well-categorized cadences. Despite this, parallelism still plays a large role in the design of a line. Rhyming lines, for example, often want to be identical to each other, even if that means an odd setting of the text. The lines in (69) are a rhyming pair: notice how the first line will have a rather normal setting.

(69) (a)



(b)



We would expect "swell enough" of (69b) to land with "-nough" on beat **1** as our stress-matching and phrasal constraints might predict. Due to parallelism, however, we get a slightly more unexpected setting because the second line wants to be like the first. Interestingly enough, "enough" could be argued to be a content or function word, so because it straddles these two worlds, it is possible to imagine it set with "-nough" on a strong beat and to find it treated as a stress-less content word. **PARALLELISM**, therefore, outranks phrasal and stress-matching constraints in favor of first having lines correspond and match with each other. There is also here an unresolved issue of cyclicity—what comes first? the parallelism or the derivation of a line?

4.4 The text-setting system as a whole

On the whole, the conflict between the constraints outlined in §4 suffice to derive most of the lines that occur in the data. There remain a few problems, which include variation (see §5) and artistic licensing. One of the major wrenches thrown in the system are the slower songs—ballads—that defy many of the text-setting goals that hold true for the medium to fast songs. This is quite understandable, though: the slower tempos of these songs makes them more susceptible to freedom than occurs in the faster, more regulated tempos. Even if there are problems with the slower songs, these problems are limited to difficulties with phrasal or line-external constraints. Stress-matching constraints still hold.

Apart from revisions of the constraints and some additions of new constraints, the text-setting system for jazz bop swing remains closely related to Hayes's OT text-setting for Western folk song. Text-setting systems, therefore, may share universal goals: match stress to strong, align phrases, avoid spilling over, make lines parallel. Even so, some of the generalizations found in jazz bop swing text-setting—like right edge stress strictness and left edge anacrusis—seem idiomatic, and I am not sure whether even Western folk song shares these traits.

5. Conclusion

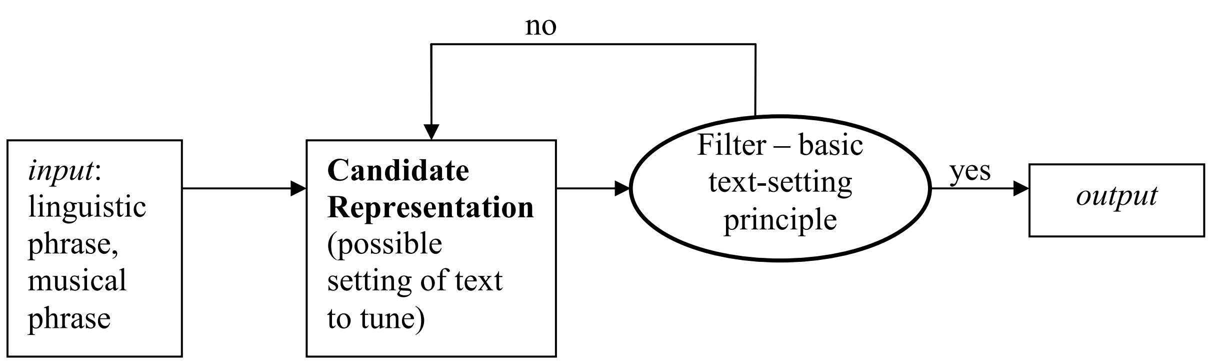
This paper applied the linguistic methods of modeling and analyzing rhythm in Western classical music and folk song to jazz bop swing. With its perceptibly different, lilting rhythm, jazz bop swing presented itself as a challenge to the previously hypothesized model of musical rhythm, first formulated by Jackendoff and Lerdahl (1983) in *The Generative Theory of Tonal Music* (*GTTM)*: the grid. Then, the question of how this rhythm of jazz bop swing corresponded to text in song required a more detailed look at the basic goals of a text-setting system. Following from the previous text-setting work by Hayes, I used an Optimality-theoretic model in which to characterize the goals of the system—constraints—and the prioritization of these goals—constraint rankings.

Jazz bop swing rhythm proved quite complex in the end, after taking into account the roles and patterns of the different instruments involved: the rhythm section instruments (r-s) and the harmonic and melodic instruments (h-m). The complex structure of jazz bop swing involves rhythmic conflict at more than one level. There is rhythmic conflict between inherent, left-edge boundary strengthening and metrically accentuated and strong beats. And there is rhythmic conflict, most notably, between the dual existence of two separate models of jazz bop swing rhythm that work in tandem to create the poly-rhythmic texture characteristic of jazz.

The goals of a text-setting system in jazz bop swing are very similar to the goals of the text-setting system in folk song verse, as demonstrated by previous work on the topic. At the most basic level, the main text-setting goal is to match linguistic stress to musical strength. Additionally, phrasal and line-external goals also contribute to judging and governing the grammaticality of the possible text-to-tune outputs given in an OT tableau. A few points did need revision, however, including a simplification of the OT-text-setting "grammar" through the use of gradient violability; moreover, the OT system proposed here reflects the relativity of the tree model as an abstract representation that allows for greater flexibility in the grammar of this art form.

The OT constraints that form the basis of modern text-setting analyses, including the one in this paper, operate at the level of text-to-tune matching: both text and tune have already been generated, and the OT-text-setting grammar developed here and in Hayes's work only judges the grammaticality of the possible text-to-tune outputs to pick the most optimal winner. The text-setting grammar at this level was deemed by Halle and Lerdahl (1993) in the first generative work on text-setting to be the "basic text-setting principle" (Halle and Lerdahl 1993), schematized below, and it has been the nature of this basic principle that researchers are primarily after.

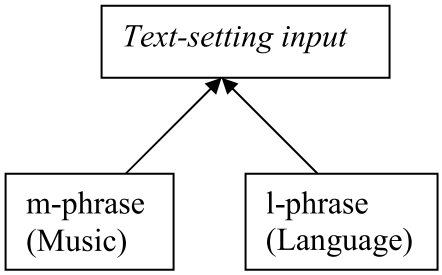
(70) adapted from Halle and Lerdahl (1993)



To briefly describe how this works, the input—a l-phrase and a m-phrase—gets thrown together in a possible candidate. This candidate is then run through the "basic text-setting principle" filter, which consists of rules or Optimality-theoretic constraints. If the candidate satisfies the filter, it proceeds to the surface output. If the candidate fails in the filter, it gets recycled back and produced as a candidate again. This flow chart is a bit outdated in that OT allows us to generate all of the possible inputs at once and then judge the best one, but the same general idea applies about the filter applies.

One important aspect of text-setting has been heretofore neglected by researchers is the question of what actually goes into the input of the text-to-tune system. There are clearly two separate entities that feed into the input to the text-setting filter: the linguistic phrase and the musical phrase.

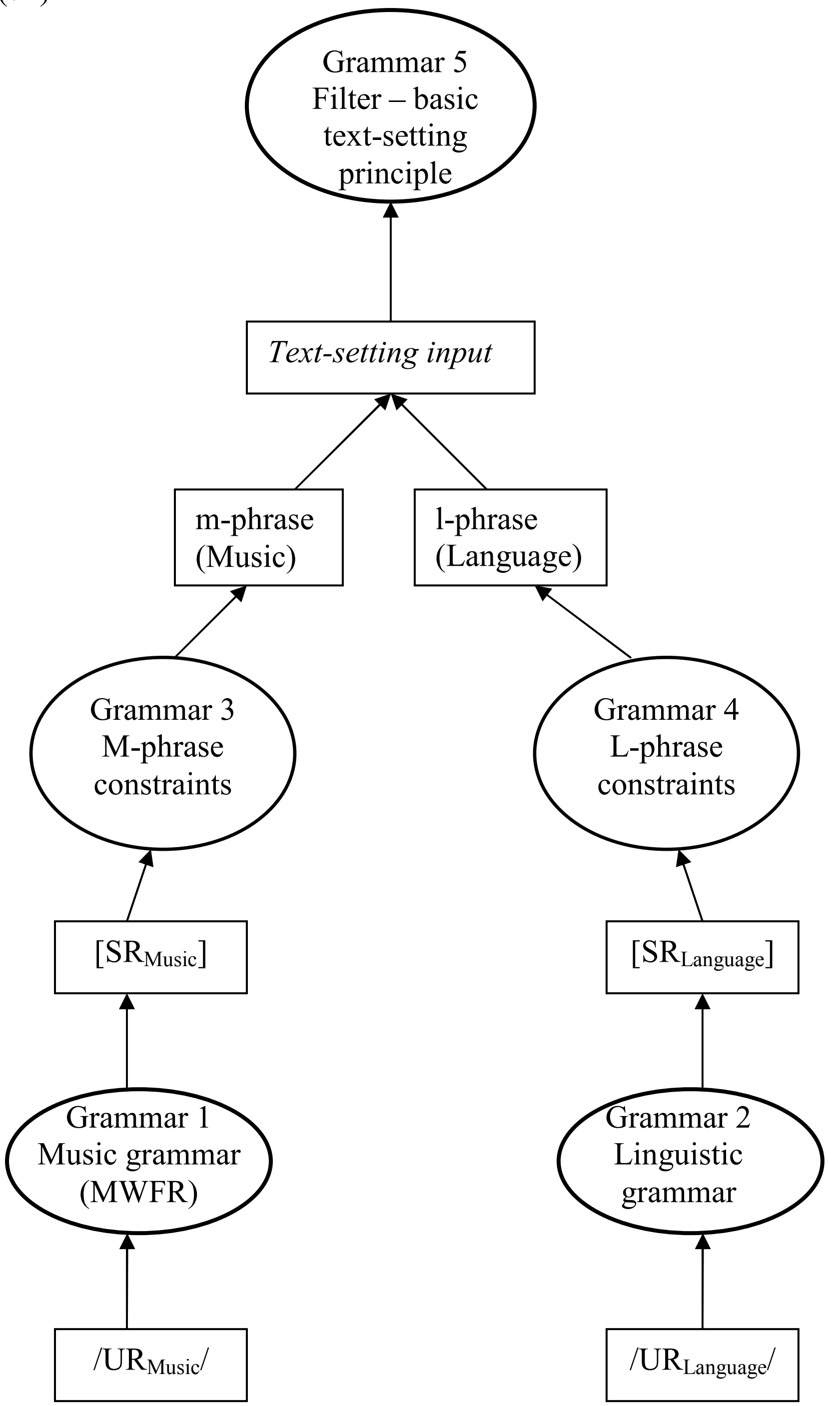
(71)



These are so-called separate entities because they each must be governed by their own grammars: one would never expect to find a piece of ungrammatical English, like *bnick, surface in a text-to-tune output. In this way, the process of having two inputs, each with its own constraint grammars, that form the text-setting input, which is then subject to a grammar of its own is akin to co-phonology theory, as applied to morpho-phonological phenomena.

Given this hypothesis, I would not be surprised to find another intermediate stage between the lowest level grammars that produce the music and the language and the highest level grammar that produces the text-setting output. The function of this intermediate stage is to govern what can possibly be chosen for input into the text-setting input. The flow chart in (72) roughly demonstrates each level of grammars.

(72)



There is preliminary evidence in favor of this intermediate stage: grammars three and four. For instance, how the system chooses the length of a m-phrase is likely to be subsumed in Grammar 3. It is consistent that m-phrases be an even number of measures long, and numbers of measures that are not a power of two are exceedingly rare in jazz bop swing. The determination of m-phrase length would not fall under the musical grammar, since the musical grammar only governs the well-formedness of the rhythmic structure.

As far as the linguistic side goes, I have already discussed the constraint **MAJOR SYNTACTIC BREAK @ EDGE**, which governs not so much the language and the music but only the content of the language. This is a constraint that could live in the realm of Grammar 4, especially because it involves developing the l-phrase prior to the text-setting input. Additionally, it is entirely possible for constraints that limit syllable count in the l-phrase input exist; for example, it is likely that the system will avoid trying to set a long sentence into an l-phrase that will be paired with an average length m-phrase—instead, the sentence will be broken down into constituents that are more conducive to setting.

Taking this co-phonology-esque approach allows exploration of other relevant topics in text-setting, such as variation. Variation is all over the place in text-setting: the same song may contain two alternative settings of the same line, an example of intra-personal variation. Variation in text-setting exists as well on an inter-personal level, with variation found from speaker to speaker. Kiparsky (2005) introduces the use of "partially ranked categorical constraints" (1 Kiparsky 2005) to model variation in text-setting, based on Arto Anttila's work with phonological and morphological variation. Anttila (2002) has extended his work to explore the possibilities of variation within co-phonologies; this approach may be useful to further research of variation in text-setting.

This paper demonstrates that jazz is at once different from and similar to Western classical music. Though most of the text-setting research has been done within the confines of

Western classical music and folk song traditions, we cannot just write off non-Western classical traditions as abnormal variants of Western classical music and thusly unworthy of detailed investigation. In fact, some of the most universal, cross-linguistic, and cross-musical tendencies may surface in non-Western classical traditions where they are otherwise masked by the institutionalization and scholarship of the Western classical tradition over the past centuries. For example, four-four (4/4), common time is not the predominant meter in the world's musics—nor is the mono-rhythmic texture that we find often in Western classical music. Even within itself, Western classical music exhibits oddities: heavy syncopation, hemiola, and three-four (3/4) or six-eight (6/8) time. The next step in rhythmic and text-setting research will have to be diversifying the data pool and exploring how the cross-musical and cross-linguistic text-setting goals found in well-behaved Western musics act in other traditions.

References

Anttila, Arto. 2002. “Morphologically conditioned phonological alternations.” *Natural Language and Linguistic Theory*. 20: 1-42.

Evans, G. Blakemore, ed. 1974. *The Riverside Shakespeare.* Boston, Houghton Mifflin.

Fitzgerald, Ella. 1960. *The* Complete *Ella in Berlin: Mack the Knife*. New York: Verve Records.

_____. 1993. *Ella Fitzgerald: First Lady of Song*. Three Disc Compilation. New York: Verve Records.

_____. 1959. *Get Happy!* New York: Verve Records.

_____. 1958. *In Rome: the Birthday Concert*. New York: Verve Records.

_____. 1961. *Sings the Harold Arlen Song Book.* New York: Verve Records.

_____. 1964. *Sings the Johnny Mercer Song Book*. New York: Verve Records.

_____. 1959. *Sings Sweet Songs for Swingers*. New York: Verve Records.

Halle, John and Fred Lerdahl. 1993. “A Generative Textsetting Model.” *Current Musicology*. 55: 3-23.

Hanson, Kristin. 2003. “Metrical Boundaries.” *MS*.

Hanson, Kristin. 1991. *Resolution in Modern Meters*. Unpublished PhD dissertation, Stanford University, Stanford, CA.

Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

_____. 1989. “The prosodic hierarchy in meter.” *Phonetics and Phonology: Volume 1, Rhythm and Meter*. ed. Paul Kiparsky and Gilbert Youmans, ed. San Diego, CA: Academic Press, Inc.

_____. 2005. “Textsetting as Constraint Conflict.” (submitted) *Typology of Poetic Forms*. ed. Jean-Louis Aroui. <http://www.linguistics.ucla.edu/people/hayes/Textsetting/HayesTextsettingAsConstraintConflict.pdf>

_____ and Abigail Kaun. 1996. “The Role of Phonological Phrasing in Sung and Chanted Verse.” *The Linguistic Review*. 13: 243-303.

______ and Margaret MacEachern. 1997. “Quatrain Form in English Folk Verse.”

Jackendoff, Ray. 1989. “A Comparison of Rhythmic Structures in Music and Language.” *Phonetics and Phonology: Volume 1, Rhythm and Meter*. ed. Paul Kiparsky and Gilbert Youmans, ed. San Diego, CA: Academic Press, Inc.

_____ and Fred Lerdahl. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Jesperson, Otto. 1933. “Notes on Metre.” in *Linguistica* (Danish original published in Oversigt (1990)).

Kager, René. 1999. *Optimality Theory*. Cambridge, UK: Cambridge University Press.

Kiparsky, Paul. 1977. “The Rhythmic Structure of English Verse.” *Linguistic Inquiry*. 8, 2: 189-247.

______. 2005. “Where Stochastic OT fails: a discrete model of metrical variation.” *Berkeley Linguistics Society*.

______. 2006. “A Modular Metrics for Folk Verse.” In B. Elan Dresher and Nila Friedberg (ed). *Formal Approaches to Poetry*, 7-49. Mouton.

Liberman, Mark and Alan Prince. 1977. “On Stress and Linguistic Rhythm.” *Linguistic Inquiry*. 8.2: 249-336.

Macy, L. 2006. “Syncopation” *Grove Music Online* ed. L. Macy (Accessed 12/11/2006), <http://www.grovemusic.com>

McCarthy, John. 2003. “OT Constraints are Categorical.” *Phonology* 20, 75-138.

Prince, Alan. 1989. “Metrical Forms.” *Phonetics and Phonology: Volume 1, Rhythm and Meter*. Paul Kiparsky and Gilbert Youmans, ed. San Diego, CA: Academic Press, Inc.

Witmer, Robert. 2006. “Fill.” *Grove Music Online* ed. L. Macy (Accessed 17/9/2006), <http://www.grovemusic.com>